



UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE ZOOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA

**O MITOGENOMA E O SATELITOMA DE *Astyanax
altiparanae* GARUTTI & BRITISKI, 2000:
CONSIDERAÇÕES INICIAIS**

Matheus Lewi Cruz Bonaccorsi de Campos

Belo Horizonte

2020

Matheus Lewi Cruz Bonaccorsi de Campos

**O MITOGENOMA E O SATELITOMA DE *Astyanax*
altiparanae GARUTTI & BRITISKI, 2000:
CONSIDERAÇÕES INICIAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Zoologia do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Zoologia.

Orientadora: Profa. Dra. Karine Frehner Kavalco

Belo Horizonte

2020

043

Campos, Matheus Lewi Cruz Bonaccorsi de.

O mitogenoma e o satelitoma de *Astyanax altiparanae* Garutti & Britiski, 2000: considerações iniciais [manuscrito] / Matheus Lewi Cruz Bonaccorsi de Campos. - 2020.

60 f. : il. ; 29,5 cm.

Orientadora: Profa. Dra. Karine Frehner Kavalco.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Zoologia.

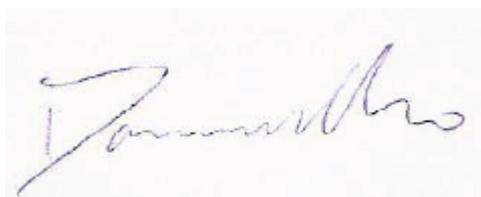
1. Zoologia. 2. DNA Mitochondrial. 3. DNA Satélite. 4. Filogenia. 5. Sequenciamento Completo do Genoma. I. Kavalco, Karine Frehner. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 591

**O mitogenoma e o satelitoma de *Astyanax altiparanae* Garutti & Britski, 2000:
considerações iniciais**

AUTOR: MATHEUS LEWI CRUZ BONACCORSI DE CAMPOS

Esta dissertação foi apresentada em sessão pública e submetida a avaliação em 24 de março de 2020, pela Banca Examinadora composta pelos seguintes membros:



Prof. Dr. Daniel Cardoso de Carvalho (Membro / PUC-MG)



Profa. Dra. Caroline Garcia (Membro – UESB)



Profa. Dra. Karine Frehner Kavalco (Orientadora – UFV)

BANCA EXAMINADORA

Dra. Karine Frehner Kavalco (Orientadora)

Universidade Federal de Viçosa

Dra. Caroline Garcia

Universidade Estadual do Sudoeste da Bahia

Dr. Daniel Cardoso de Carvalho

Pontifícia Universidade Católica de Minas Gerais

Dr. Almir Rogério Pepato (suplente)

Universidade Federal de Minas Gerais

DEDICATÓRIA

À minha tia Maria Inês, que me fez apaixonar pelas Ciências Naturais enquanto criança e alimentou a curiosidade que me fez ser Cientista. “Quanto mais esdrúxulo, melhor!”, dizia ela.

AGRADECIMENTOS

Primeiramente, eu gostaria de agradecer aos meus pais acadêmicos, orientadores e grandes amigos, os professores Karine Kavalco e Rubens Pasa, que me guiaram não só na ciência, desde o começo da graduação, mas também na vida, quando minha mente ficava deveras confusa;

Gostaria de agradecer também aos professores Caroline Garcia e Daniel Cardoso, por terem aceitado a tarefa de compor minha banca examinadora, contribuindo assim com a minha dissertação e formação, e aos professores John Seymour (Pat) Heslop-Harrison e Trude Schwarzacher, da *University of Leicester*, pela inestimável parceria, tanto intelectual quanto financeira;

Aos amigos de laboratório, em especial à Dinaíza Reis e à Snaydia Resende, que sempre com paciência e disposição, me cederam seu tempo para ajudar nos trabalhos e ouvir minhas lamúrias. Não sei o que seria de mim sem essas duas;

Ao laboratório em si, que com os outros alunos também grandes amigos, se tornou uma segunda casa, tanto nas saídas de campo (agradecimento especial ao Iuri Silva, inestimável e excelente companheiro de coletas), quanto nas bancadas abrindo peixes e também nos vários dias de churrasco;

Ao Fabiano Menegídio, que me apresentou a Bioinformática e dispôs de todo o tempo do mundo, sempre com entusiasmo e paciência, para me orientar nessa obscura área;

Aos eternos amigos de república, especialmente ao Francisco Sassi, ao Vitor Augusto e à Mayumi Furuya, que, cada um ao seu modo, me ensinaram valiosas lições acadêmicas e de vida, fazendo dos últimos 3 anos os mais divertidos;

Aos inestimáveis amigos do mestrado, Amanda, Bárbara, Caroliny, Júlia, Lucas, Pedro e Osvaldo, que não só ajudaram academicamente, mas também fizeram da tão temida e pesada pós-graduação, uma jornada leve e descontraída;

À minha família, mãe e avó, que sendo as duas mulheres mais incríveis e guerreiras desse mundo, me motivam à tentar voar sempre alto e mesmo quando começo a cair, são elas que me levantam de novo. Ao meu pai, que mesmo sem entender muito

bem minhas escolhas, também sempre me apoiou em cada uma delas e me ensinou que se “for pra trabalhar, é pra trabalhar de verdade e fazer bem feito”;

À Universidade Federal de Viçosa e à Universidade Federal de Minas Gerais, por toda a infraestrutura que permitiu a realização deste trabalho e pelos professores, todos de indubitável excelência, que com muita admiração tive o prazer de conhecer e que contribuíram tanto com os conhecimentos que vou carregar pelo resto da vida;

À FAPEMIG e ao CNPq, respectivamente pela bolsa concedida e pelos auxílios financeiros que foram fundamentais para a execução da pesquisa ao longo do mestrado;

À Alexandra Elbakyan, responsável por “remover todas as barreiras no caminho da ciência” graças ao SciHUB. É possível contar nos dedos quantos dos artigos aqui citados eram genuinamente gratuitos, todos os demais dificilmente seriam acessíveis se não fosse por ela;

Por fim, dedico um agradecimento especial à Ravena, minha gata, pois não importa o quão ruim esteja o dia, sempre que esfrego e afago aquela bola de pelos, tudo melhora.

A todos vocês, meus eternos agradecimentos!

“Seja qual for o rumo que tomarmos, nosso destino está indissolúvelmente ligado à ciência.”

“A ciência é muito mais do que um corpo de conhecimento, é uma forma de pensar”

- Carl Sagan

SUMÁRIO

RESUMO	01
ABSTRACT	02
1. INTRODUÇÃO	03
1.2 OBJETIVOS	11
2. CAPÍTULO 1: MONTAGEM DO GENOMA MITOCONDRIAL COMPLETO DE <i>Astyanax altiparanae</i> Garutti & Britski, 2000 (CHARACIFORMES, CHARACIDAE).....	12
2.2 INTRODUÇÃO	13
2.3 MATERIAIS E MÉTODOS	14
2.4 RESULTADOS	15
2.5 DISCUSSÃO E CONCLUSÕES	19
2.6 MATERIAL SUPLEMENTAR	21
3. CAPÍTULO 2: IDENTIFICAÇÃO <i>DE NOVO</i> E INVESTIGAÇÃO <i>IN SILICO</i> DO SATELITOMA DE <i>Astyanax altiparanae</i> (CHARACIFORMES, CHARACIDAE).....	23
3.2 INTRODUÇÃO	24
3.3 MATERIAIS E MÉTODOS	27
3.4 RESULTADOS E DISCUSSÕES	29
3.5 CONCLUSÕES	40
3.6 AGRADECIMENTOS	40
3.7 MATERIAL SUPLEMENTAR	41
4. CONSIDERAÇÕES FINAIS	43
5. LISTA DE ABREVIACÕES	45
6. REFERÊNCIAS BIBLIOGRÁFICAS	47

RESUMO

Dentre os peixes neotropicais, o gênero *Astyanax* é considerado como um dos mais diversos e abundantes. São peixes conhecidos popularmente como lambaris ou piabas e apesar de serem amplamente estudados, suas relações filogenéticas ainda são incertas, em função de uma taxonomia complexa repleta de espécies crípticas e complexos de espécies. Há algumas décadas, estudos com citogenética clássica e molecular têm fornecido dados valiosos sobre a dinâmica cromossômica do gênero, ajudando à identificar os complexos de espécie e à traçar suas histórias evolutivas. Seguindo os avanços nas tecnologias de sequenciamento de genomas e na biologia computacional, uma nova área tem crescido, a Ciência de Dados Genômicos, fornecendo uma infinidade de novas ferramentas e abordagens para estudos genéticos, ecológicos e evolutivos em diversos organismos. Nesta dissertação, exploramos algumas dessas ferramentas no genoma de *Astyanax altiparanae*, que, diferentemente de outras espécies do gênero, é cromossomicamente pouco diversa, onde os marcadores citogenéticos atuais não se mostram muito resolutivos. Com os dados genômicos obtidos por sequenciamento de nova geração (NGS) e utilizando de uma série de ferramentas de bioinformática, objetivamos realizar a montagem de um mitogenoma completo para *A. altiparanae* e analisar o satelitoma desta espécie. O mitogenoma obtido foi descrito e foi feita a anotação funcional do mesmo. Construímos uma filogenia a partir desse mitogenoma em conjunto com os de outros Characiformes disponíveis na literatura e exploramos o satelitoma de *A. altiparanae*, onde descobrimos e classificamos 32 sequências de DNAs satélites promissores para a espécie. Com base nos resultados obtidos, concluímos que a filogenia baseada em genomas mitocondriais completos dispõe a espécie *A. altiparanae* como grupo irmão de *A. mexicanus* e reforça a ideia de uma condição parafilética para o gênero. Além disso, das 32 sequências de elementos repetitivos identificadas, sugerimos três como possíveis novos marcadores citogenéticos para *A. altiparanae* e para *Astyanax* como um todo. Por fim, demonstramos que as abordagens *in silico* se inserem como um complemento forte para as técnicas clássicas de citogenética e FISH, sendo desde já imprescindíveis uma vez que, ao suprir algumas das limitações da citogenética clássica e da FISH, fornecem novos pontos de partida para melhor entender os mecanismos genéticos e cromossômicos por detrás dos padrões evolutivos dos *Astyanax*, ajudando assim, numa reconstrução mais precisa da história do gênero.

ABSTRACT

Among neotropical fish, the genus *Astyanax* is considered one of the most diverse and abundant. These fish are popularly known as “tetras” or “piabas” and despite being widely studied, their phylogenetic relationships are still uncertain, due to a complex taxonomy filled with cryptic species and species complex. For some decades, classical cytogenetics and molecular studies have provided valuable data on the chromosomal use of the genus, helping to identify species complexes and track their evolutionary histories. Following advances in genome sequencing technologies and computational biology, a new area has grown, Genomic Data Science, providing a wilderness of new tools and approaches for genetic, ecological and evolutionary studies in diverse organisms. In this dissertation, we explore some of these tools in the *Astyanax altiparanae* genome, which, unlike other species of the genus, has poor chromosomal diversity, where current cytogenetic markers can be less effective. With the genomic data obtained by Next-Generation Sequencing (NGS) and using a series of bioinformatics tools we aim to carry out the assembly of a complete mitogenome for *A. altiparanae* and analyze the satellite of this species. The mitogenome obtained was described and its functional annotation was made. We built a phylogeny from this mitogenome together with those of other characiforms available in the literature and explored the *A. altiparanae* satellite, where we discovered and classified 32 promising satellite DNA sequences for the species. Based on the results obtained, we conclude that phylogeny based on complete mitochondrial genomes makes the *A. altiparanae* species the sister group of *A. mexicanus* and reinforces the idea of a paraphyletic condition for the genus. In addition, of the 32 sequences of repetitive elements identified, we suggest three as possible new cytogenetic markers for *A. altiparanae* and for the genus as a whole. Finally, we demonstrate that *in silico* approaches are a strong complement to the classic techniques of cytogenetics and FISH, being essential since, by addressing some of the limitations of these two techniques, they provide new starting points to better understand the genetic and chromosomal mechanisms behind the evolutionary patterns of *Astyanax*, thus helping in a more accurate reconstruction of the history of the genre.

1. INTRODUÇÃO:

Os “Peixes” configuram um grupo polifilético de animais vertebrados que se distribuem por águas doces e marinhas de quase todo o globo Terrestre (Helfman *et al.*, 2009). Esses animais correspondem a mais do que a metade de todas as espécies de vertebrados vivos atualmente. Deste modo, também é alto o número de espécies no grupo, cuja diversidade está partilhada em aproximadamente 540 famílias de 85 ordens. Os trópicos abrigam o maior número de espécies dessas famílias, onde dois terços delas são peixes que habitam águas doces (Nelson *et al.*, 2016).

Dentre as famílias existentes de peixes, Characidae é uma das maiores e mais complexas: é considerada a quarta mais diversa do mundo e a mais diversa quando se trata das famílias Neotropicais, onde possui mais de 1000 espécies distribuídas entre o Norte da Argentina e o Sul dos EUA (Mirande, 2009 e 2010). Grande parte dos Characidae é morfológicamente bem conservada, entretanto, diversas fontes importantes de variação morfológica, relacionada principalmente à alimentação, à resistência corporal aos diferentes fluxos de águas (com e sem corredeiras) (Breda *et al.*, 2005) e até eventos de miniaturização, resultam em uma série de caracteres altamente homoplásticos no grupo (Mirande, 2010).

As relações filogenéticas dos Characidae, tanto as internas quanto as com os demais Characiformes, são instáveis (Mirande, 2018) e têm sido questionadas e discutidas há décadas na literatura, em função de dúvidas recorrentes quanto à sua composição e monofiletismo (Lucena, 1993; Oliveira *et al.*, 2011; Mirande, 2018). Uma definição mais restritiva para Characidae foi proposta por Oliveira e coautores em 2011, onde a família foi fortemente recuperada como monofilética e caracterizada morfológicamente pela ausência de um osso supra-orbital em seus integrantes (Malabarba & Weitzman, 2003; Oliveira *et al.*, 2011).

Essa e outras sinapomorfias utilizadas para alcançar estabilidade monofilética em Characidae, embora sejam consideradas uma solução compatível com abordagens moleculares e morfológicas disponíveis na literatura, ainda não resolvem totalmente as relações de seus ramos internos (Mirande, 2018). Um dos motivos disso é a presença de mais de 80 gêneros considerados como *incertae sedis* (do latim “com posição incerta”) nesta família (Lima *et al.*, 2003; Oliveira *et al.*, 2011; Mirande, 2018).

Um desses gêneros é o *Astyanax*. Na mitologia grega, *Astyanax* significava “Lorde da cidade”, foi o apelido dado pelos Troianos à Scamandrius, príncipe filho de Hector e Andrômaca, durante a Guerra de Tróia (Homer, 1924). Scamandrius era uma homenagem ao rio Scamander, ao noroeste de Anatólia (Turquia moderna), e ao Deus-rio de mesmo nome (Homer, 1924; Koerber, 2011; Rose & March, 2015; *Encyclopædia Britannica*, 2018).

Como gênero de peixes, *Astyanax* foi descrito por Baird & Girard em 1854 com base no *Astyanax argentatus*. Eles incluíram como “Characini” peixes caracterizados pela barbatana dorsal acima das barbatanas pélvicas; linha lateral não serrilhada; presença de uma nadadeira adiposa; fileiras duplas de dentes em ambas as mandíbulas achatadas; sem dentes caninos ou palatinos e com escamas grandes (Garavello & Sampaio, 2010).

Entre 1921 e 1927, em sua revisão do gênero, Eigenmann incluiu uma série de novos caracteres morfológicos para definir o grupo, como o comprimento máximo entre 150 e 200mm e linha lateral completa. Isso abrangeu 74 novas espécies e subespécies distribuídas do Rio Grande, na fronteira entre México e EUA, até o sul da Argentina, Uruguai e Brasil (Eigenmann, 1921 e 1927; Garavello & Sampaio, 2010).

Atualmente o gênero compreende cerca de 170 espécies de peixes pequenos conhecidos popularmente como piabas, lambaris ou “*tetra fishes*” (Eschmeyer, 2019). É o gênero mais rico em espécies dentro de Characidae, o mais diverso da região Neotropical (Orsi, 2004), sendo também um dos mais abundantes na América do Sul (Eigenmann, 1921).

A alusão do gênero *Astyanax* ao Deus-rio é compreensível dada a tamanha diversidade encontrada no grupo. Essa diversidade é, em partes, reflexo do hábito de vida e da história evolutiva destes peixes, que se dá em populações isoladas ao longo de riachos de cabeceira e pequenos tributários de corredeiras aos quais eles normalmente estão restritos (Garutti & Britski, 2000).

Os riachos de cabeceira contêm, desproporcionalmente, um grande número de espécies em geral (Nelson *et al.*, 2016) e naturalmente, a ictiofauna neotropical como um todo é caracterizada por um alto grau de endemismos (Albert e Reis, 2011). Isso acontece, pois, populações isoladas geralmente exibem uma grande variabilidade quando se trata de características morfológicas e genéticas (Badyaey *et al.*, 2000).

Ainda se tratando da complexa diversidade dos *Astyanax*, um de seus efeitos é a existência de espécies consideradas crípticas e, conseqüentemente, a incerteza que essas espécies apresentam quanto às suas relações filogenéticas. Como exemplo disso pode-se citar dentro deste gênero, a construção artificial de três principais complexos de espécies: *A. bimaculatus* (Garutti, 1995), *A. fasciatus* (Justi, 1993) e *A. scabripinnis* (Moreira-Filho & Bertollo, 1991). Estes diferentes complexos seriam uma forma intuitiva de agrupar conjuntos de espécies que compartilham características em comum, porém, suas filogenias não corroboram com tais agrupamentos.

O gênero como um todo é considerado polifilético e suas relações filogenéticas são incertas (Oliveira *et al.*, 2011; Schmitter-Soto, 2016). Sendo assim, mesmo que filogenias recentes aloquem os *Astyanax* dentro da subfamília Stethaprioninae (Mirande, 2018), alguns autores consideram o gênero como *Incertae sedis*, por não apresentar evidências robustas de monofiletismo, (Oliveira *et al.*, 2011; Schmitter-Soto, 2016; Mirande, 2018).

Uma outra questão que envolve o gênero, dado o difícil reconhecimento e identificação de suas espécies, é a inconsistência que pode ser encontrada na literatura quando a taxonomia leva à classificação errônea daquelas que são morfologicamente parecidas ou crípticas. A atual espécie *A. fasciatus* pode ilustrar bem isso: *A. fasciatus* foi descrita originalmente por Cuvier em 1819 como *Chalceus fasciatus*. Os peixes que remetiam à *Chalceus fasciatus* passaram a ser considerados como *Tetragonopterus* por Steidachner em 1876, onde *Tetragonopterus aeneus* Hensel (1870) foi um sinônimo júnior da atual *A. fasciatus* Fowler, 1948 (Melo & Buckup, 2006).

Posteriormente, Eigenmann transferiu a maior parte dos *Tetragonopterus* para o gênero *Astyanax*, incluindo o nome júnior de *fasciatus*. Foi quando finalmente *A. fasciatus* foi redefinido por ele, entretanto, como um táxon composto por cinco subespécies, com ocorrência geral pelos rios do Brasil (Eigenmann, 1921).

A descrição original de *A. fasciatus* feita por Eigenmann foi baseada em um espécime oriundo da bacia do rio São Francisco, ainda assim ele atribuiu sua distribuição a diferentes bacias pelo país (Eigenman, 1921). Em consequência disso, alguns trabalhos têm questionado a validade dessa descrição, cujas definições são tão amplas que podem abranger outras espécies de *Astyanax*, tanto que existem *A. fasciatus* sendo relatados em

rios do México e Argentina quando deveriam ser restritos ao rio São Francisco no Brasil (Melo & Buckup, 2006).

Dos complexos de *Astyanax* citados anteriormente, pode-se destacar o *A. bimaculatus*, que apresenta características bem mais conservadas que os demais complexos (Domingues *et al.*, 2007; Pazza *et al.*, 2018). Proposto por Garutti em 1995, os *A. bimaculatus* são atualmente representados pela espécie homônima em conjunto com outras como *A. altiparanae*, *A. lacustris*, *A. assuncionensis* e *A. abramis*. Estas espécies vivem entre a vegetação submersa de águas calmas e menos turbulentas. Possuem uma alimentação versátil e comportamento oportunista, sendo distribuídas principalmente pelas bacias do alto rio Paraná, Paraguai, Iguaçu e São Francisco (Shibatta *et al.*, 2002; Domingues *et al.*, 2007; Pazza *et al.*, 2018).

A maior parte das espécies animais e vegetais possuem um número cromossômico ou citótipo característico, mas quando se trata de peixes, muitas variações podem ser encontradas, em função de rearranjos numéricos, estruturais e da presença de cromossomos supranumerários (Pazza *et al.*, 2007). A existência de uma grande diversidade cariotípica nas populações de *Astyanax*, aliada à sua abundância e riqueza, tornam o gênero um importante modelo para se estudar a diversificação evolutiva. Os *Astyanax*, em conjunto com o seu hábitat, são então alvos de diversos tipos de estudos ecológicos e genéticos que buscam elucidar seus complexos padrões evolutivos (Badyaev *et al.*, 2000).

Tais estudos são realizados geralmente a partir da Citogenética, que engloba a investigação dos cromossomos, suas formas e arranjos (isolados ou em conjunto), suas funções, origens, variações e evolução (Guerra, 1988); da Filogeografia e da Biogeografia, que têm fornecido muitos dados sobre a natureza demográfica e histórica da evolução intraespecífica de muitas espécies (Avice, 2009).

Os estudos cariotípicos realizados em *Astyanax* vêm fornecendo dados significativos para entender como o isolamento repercute nas suas populações (Shibatta & Artoni, 2005). Enquanto os *A. fasciatus* têm demonstrado um número cromossômico variável de $2n=45, 46, 47, 48$ e 50 (Pazza, 2006; Ferreira-Neto, 2012), os *A. bimaculatus*, em especial os *A. altiparanae* aqui estudados, possuem um número diploide constante em suas populações, sendo $2n=50$ cromossomos (Kavalco *et al.*, 2011; Fernandes *et al.*, 2014).

Esse número diploide conservado dos *A. altiparanae* é considerado simplesiomórfico dentro dos *Astyanax* (Kavalco *et al.*, 2011). Ainda assim, existem variações cariotípicas na espécie, onde são encontradas diferenças em sua fórmula cariotípica, no número fundamental de braços e na simetria geral dos cariótipos. É possível também observar citótipos distintos e estruturação populacional nesses peixes em diferentes bacias e sub-bacias (Kavalco *et al.*, 2011; Fernandes *et al.*, 2014).

Esses dados citogenéticos, em conjunto com os moleculares e os caracteres morfológicos compartilhados pelos *A. altiparanae*, como a própria mancha umeral oval e o rabo amarelo característicos, sugerem que pelo menos o complexo *A. bimaculatus* seja monofilético (Pazza *et al.* 2018), com divergência relativamente recente, onde seria composto por dois principais clados: *A. altiparanae* e *A. bimaculatus* (Pazza & Kavalco, 2007; Mello, R. 2009; Kavalco *et al.* 2011; Fernandes *et al.* 2014; Pazza *et al.* 2018).

Dentre as várias técnicas existentes dentro da Citogenética, são os trabalhos realizados com Híbridões Fluorescentes *in situ* (FISH), utilizando de sondas de DNA Ribossômico (rDNA) 5S e 18S, juntamente com o satDNA As-51, que têm fornecido dados promissores sobre a história evolutiva dos *Astyanax* (Abel *et al.* 2006; Pazza *et al.*, 2018). A combinação desses marcadores é interessante e promissora uma vez que os rDNAs 5S e 18S constituem sequências repetitivas altamente conservadas de pares de base, ao passo que o DNA satélite repetitivo denominado As-51, é menos conservado e mais plástico, devido à sua semelhança com elementos transponíveis (Martins & Galetti, 1999; Hatanaka & Galetti, 2004).

As técnicas de FISH permitem determinar e localizar fisicamente nos cromossomos a presença de sequências de interesse (que podem ser utilizadas como marcadores citogenéticos), por meio de sondas fluorescentes de DNAs complementares a essas sequências (Schwarzacher, 2003). Os fluorocromos utilizados na marcação das sondas são moléculas que ao serem excitadas por luzes de um determinado comprimento de onda, emitem luz de um comprimento menor (Schwarzacher, 2003). A luz emitida pelos fluorocromos pode então ser captada por microscópios de epifluorescência que tornam possível a observação, registro fotográfico, quantificação e comparação dos diferentes marcadores citogenéticos. Essas abordagens englobadas pela Citogenética Molecular, em conjunto com a Clássica, permitem a detecção, comparação e estudo de regiões homeólogas entre os cromossomos de diferentes espécies e populações, auxiliando na determinação de suas relações filogenéticas (Pazza *et al.*, 2018).

O genoma de um eucarioto geralmente é composto tanto por sequências de DNA únicas, quanto por sequências repetitivas como os rDNAs e o satélite As-51 mencionados anteriormente (López-Flores & Garrido-Ramos, 2012). A fração repetitiva do DNA pode ser organizada em elementos moderadamente ou altamente repetitivos, que se dispõem em conjunto ou dispersos pelo genoma, onde formam famílias de diferentes classes gênicas como elementos transponíveis, DNAs ribossômicos, pseudogenes, repetições em tandem ou DNAs satélites, constituindo o que é denominado de *Repeatome* (López-Flores & Garrido-Ramos, 2012; Ruiz-Ruano *et al.*, 2016).

DNAs satélites (satDNAs), num geral, são frações não codificantes e extremamente dinâmicas do genoma que se encontram repetidas em tandem ou dispersas, sendo um dos mais abundantes componentes do Repetoma (do inglês *Repeatome*) e dos genomas dos eucariotos (Plohl *et al.*, 2012). Um único genoma eucarioto pode carregar centenas de famílias de satDNAs, com diferentes localizações no genoma, cada uma com suas próprias dinâmicas e taxas de evolução (Ruiz-Ruano *et al.*, 2016). Todas as espécies vivas contêm então uma biblioteca de satDNAs, cujo conjunto pode ser chamado de *Satellitome* (Ruiz-Ruano *et al.*, 2016; Duílio *et al.*, 2017). O satelitoma de uma espécie pode ou não ser parcialmente ou totalmente compartilhado com outras espécies e populações proximamente relacionadas (Ruiz-Ruano *et al.*, 2016; Rodrigues *et al.*, 2019), seguindo o que é conhecido como “hipótese da biblioteca” (Fry & Salser, 1977).

Essas sequências repetitivas mudam rapidamente ao longo da evolução dos organismos, onde se tornam excelentes marcadores de segmentos cromossômicos que permitem rastrear os rearranjos cromossômicos (Moscone *et al.*, 2007). Apesar de fornecerem diversos dados valiosos, estudos genômicos e citogenéticos sobre os satelitomas ainda são escassos, principalmente se comparados a famílias multigênicas como a dos DNAs ribossômicos (Rebordinos *et al.*, 2013). Isso talvez se deva aos laboriosos e complexos métodos de obtenção desse tipo de DNA nos diversos organismos (Heslop-Harrison, 2000; Vicari *et al.*, 2010; Mehrotra *et al.*, 2014; Pamponét *et al.*, 2019).

Os métodos mais tradicionais para o isolamento de sequências de DNAs satélites em peixes Actinopterygii, envolviam o uso de enzimas de restrição específicas, reassociação cinética com DNA C0t-1 (DNA enriquecido com sequências altamente e moderadamente repetitivas) e a microdissecção de cromossomos submetidos à bandamento-C, seguidos da amplificação de sequências heterocromáticas via DOP-PCR

(reação em cadeia da polimerase iniciada com oligonucleotídeos degenerados) (Telenius *et al.*, 1992; Ferreira & Martins *et al.*, 2008; Vicari *et al.*, 2010).

Nos últimos anos, o sequenciamento de nova geração (NGS) a custos mais baixos, em conjunto com ferramentas de bioinformática mais acessíveis, têm facilitado a identificação, montagem, caracterização e comparação das sequências repetitivas (e até mesmo de genomas completos) entre as espécies (Novák *et al.*, 2013; Afgan *et al.*, 2018; Rodrigues *et al.*, 2019).

Quando se fala de uma bioinformática mais acessível, todo o destaque vai para a plataforma pública Galaxy (<http://usegalaxy.eu/> - Afgan *et al.*, 2018). Iniciada em 2005, a plataforma Galaxy é um sistema para integração e manipulação de sequências de DNA, alinhamentos, montagens e anotações genômicas. Atualmente, o Galaxy está disponível em três servidores diferentes, cada um dispo de um conjunto distinto de ferramentas, um servidor europeu (<http://usegalaxy.eu/>), um na Austrália (<https://usegalaxy.org.au/>) e um americano (<https://usegalaxy.org>). De acordo com seus desenvolvedores, o sistema habilitou biólogos sem treinamento em programação e administração de sistemas de computadores, à desenvolverem análises computacionais robustas via web (Afgan *et al.*, 2018).

Diversos tipos de análises podem ser performadas rodando diferentes séries de ferramentas sucessivas (que podem ser automatizadas através de workflows construídos pelo próprio usuário) dentro da plataforma, a partir de qualquer navegador de internet, onde o Galaxy preserva a procedência, inputs, outputs e o referencial bibliográfico de cada etapa realizada (Afgan *et al.*, 2018). Para estudos de satelitoma, por exemplo, é possível usar uma instância alternativa do sistema Galaxy que é o RepeatExplorer (<https://repeatexplorer-elixir.cerit-sc.cz/> - Novák *et al.*, 2013).

O servidor do RepeatExplorer fornece uma ferramenta homônima cuja *pipeline* (instruções segmentadas) computacional utiliza de *reads* (leituras) gerados por NGS de *shotgun* para identificação de sequências repetitivas (Novák *et al.*, 2013; Lower *et al.*, 2018). Essa abordagem já permitiu não só a busca por sequências que podem ser utilizadas como sondas e marcadores citogenéticos na FISH em diversos organismos (de plantas dos gêneros *Quercus* e *Passiflora* (Pamponét *et al.*, 2019; Mascagni *et al.*, 2020) à animais-modelo como moscas (Silva *et al.*, 2019), gafanhotos (Ferretti *et al.*, 2019) e peixes (Duílio *et al.*, 2017), como também permitiu a classificação e o estudo comparativo

in silico de contigs de DNAs repetitivos que são invisíveis nas hibridações (Ruiz-Ruano *et al.*, 2016; Duílio *et al.*, 2017).

A plataforma Galaxy, incluindo suas instâncias, também se atém a questões delicadas de replicabilidade e reprodutibilidade das análises de dados genômicos (Goecks *et al.*, 2010), onde ela não só possui integração com alguns bancos de dados genômicos como o GenBank, como também permite que suas ferramentas, workflows e *pipelines* sejam facilmente exportados e compartilhados. Deste modo, o Galaxy tem sido utilizado para realizar diferentes tipos de pesquisas genômicas de ponta, indo da montagem e anotação estrutural e funcional de genomas (*de novo* ou com genoma de referência) à investigações de epigenômica e de interações genoma-ambiente (Kikuchi *et al.*, 2010; Peleg *et al.*, 2010).

O sequenciamento de genoma completo em conjunto com o Galaxy, podem então ser utilizados como uma nova e alternativa abordagem para se estudar a dinâmica evolutiva dos *Astyanax*. As ferramentas de bioinformática disponíveis na plataforma auxiliam no estudo *in silico* de sequências genômicas obtidas via NGS. Essas sequências aliadas às técnicas de FISH, podem se tornar marcadores citogenéticos bem mais eficientes, uma vez que já foram previamente identificadas e estudadas (Ferretti *et al.*, 2019; Rodrigues *et al.*, 2019).

Dentro desses dados de genoma completo, as sequências que compõem todo o DNA mitocondrial (mtDNA) também podem ser recuperadas. Tal qual os marcadores citogenéticos que ajudam à traçar filogenias cromossômicas, o mitogenoma à parte se mostra igualmente promissor quando se tratam de análises filogenéticas moleculares, estudos populacionais e de estruturação genética das diferentes populações (Avise *et al.*, 1987). Historicamente, genes mitocondriais têm sido muito utilizados em inúmeros organismos, sequências como COI, ATPase e Cyt b, são promissoras para se estudar filogenias profundas e divergências antigas entre os grupos animais (Curole & Kocher, 1999), entretanto, abordagens que utilizam do genoma mitocondrial completo ainda são relativamente recentes (Smith, 2015).

Mesmo que centenas de artigos com descrições de genoma mitocondrial tenham sido produzidos nos últimos anos e apesar de ser o genoma mais sequenciado atualmente (Smith, 2015), poucos mitocondriais completos estão disponíveis para o gênero *Astyanax*, sendo que a única espécie cujos dados genômicos são bem explorados é o *A. mexicanus*,

por ser um importante organismo-modelo em estudos evolutivos e de desenvolvimento embrionário (Borowski, 2008; Barreto *et al.*, 2017).

1.2 OBJETIVOS:

Em função do contexto apresentado na introdução, o presente trabalho teve como objetivo geral, explorar os dados oriundos do sequenciamento do genoma completo de *Astyanax altiparanae*.

1.2.1 Objetivos Específicos:

- Extrair dos dados de NGS o genoma mitocondrial completo da espécie;
- Fazer a anotação funcional desse mitogenoma;
- Descrever e comparar o mitogenoma montado de *A. altiparanae* com outros mitogenomas de peixes disponíveis na literatura;
- Realizar, *in silico* (ou seja, computacionalmente), a prospecção de sequências genômicas que possam se tornar marcadores citogenéticos úteis para estudos cromossômicos em *A. altiparanae* e em *Astyanax* como um todo.
- Levantar dados que contribuam não só para a resolução da complexa filogenia de *Astyanax*, como também para um melhor entendimento dos mecanismos genéticos e cromossômicos por trás dos padrões evolutivos e da diversidade cromossômica encontrados no gênero.

2. CAPÍTULO 1:

MONTAGEM DO GENOMA MITOCONDRIAL COMPLETO DE *Astyanax altiparanae* Garutti & Britski, 2000 (CHARACIFORMES, CHARACIDAE)

Matheus Lewi Bonaccorsi de Campos^{1,2,5}, Fabiano Bezerra Menegídio³, John Seymour (Pat) Heslop-Harrison⁴, Trude Schwarzacher⁴, Rubens Pasa^{2,5}, Karine Frehner Kavalco^{2,5}

¹ Programa de Pós-Graduação em Zoologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

² Laboratório de Genética Ecológica e Evolutiva, Universidade Federal de Viçosa, campus Rio Paranaíba, Rio Paranaíba, MG, Brasil.

³ Núcleo Integrado de Biotecnologia, Universidade de Mogi das Cruzes, Brasil.

⁴ Department of Genetics and Genome Biology, University of Leicester, United Kingdom.

⁵ Laboratório de Bioinformática e Genômica, Universidade Federal de Viçosa, campus Rio Paranaíba, Brasil.

2.1 Resumo:

Astyanax é um dos gêneros de peixes mais abundantes da América do Sul. Apesar de ser amplamente estudado, suas relações filogenéticas ainda são incertas, em função de uma taxonomia complexa, que inclui grupos de espécies crípticas. Nesse caso, filogenias mitocondriais podem ser úteis não só quando dados morfológicos e moleculares não são resolutivos, mas também para uma identificação mais precisa dos táxons. A partir da incerteza filogenética existente nos *Astyanax* e da escassez de trabalhos com sequenciamento de genoma mitocondrial completo no gênero, apresentamos aqui o mitogenoma completo de *Astyanax altiparanae*, com uma breve filogenia desse mitogenoma com os de outros Characidae. O comprimento do mitogenoma é de 16.730pb, contendo 13 genes codificadores de proteínas, 2 RNAs ribossômicos, 22 RNAs transportadores e uma região de controle com 1.067 pb (D-loop). O mitogenoma obtido segue a conformação típica dos vertebrados onde todos os genes são codificados na cadeia pesada, exceto o gene ND6 e oito tRNAs. Em nossas análises filogenéticas, *A. altiparanae* aparece como grupo irmão de *A. mexicanus*, sendo o gênero *Astyanax* paralifilético.

2.2 Introdução:

Conhecidos popularmente como lambaris, piabas ou *tetras*, os *Astyanax* compreendem um dos gêneros de peixes mais abundantes da América do Sul (Eigenmann, 1921). Com cerca de 170 espécies (Eschmeyer, 2019) distribuídas entre o sul dos EUA e o norte da Argentina (Ornelas-Garcia *et al.*, 2008), este é o gênero mais diverso da região Neotropical (Orsi, 2004).

Compartilhando uma série de traços morfológicos homoplásticos (Ornelas-Garcia *et al.*, 2008), os *Astyanax* são de difícil identificação e reconhecimento, onde apresentam espécies crípticas (Melo & Buckup, 2006; Pansonato-Alves *et al.* 2013) e complexos de espécies. Um destes complexos é o grupo *A. bimaculatus* (Garutti, 1995), que engloba os lambaris de rabo amarelo com mancha umeral oval, como a própria espécie *A. bimaculatus*, em conjunto com *A. altiparanae*, *A. lacustris*, *A. abramis*, entre outras (Domingues *et al.*, 2007; Pazza *et al.*, 2018).

Dados citogenéticos, moleculares e morfológicos sugerem que o complexo *A. bimaculatus* seja monofilético, formado por dois principais clados: *A. altiparanae* e *A. bimaculatus* (Fernandes *et al.*, 2014; Pazza *et al.*, 2018), entretanto, em função de filogenias morfológicas e moleculares abrangendo outros complexos, como *A. fasciatus* (Justi, 1993) e *A. scabripinnis* (Moreira-Filho & Bertollo, 1991), o gênero como um todo é considerado polifilético e suas relações filogenéticas são incertas (Oliveira *et al.*, 2011; Schmitter-Soto, 2016).

Ainda que filogenias recentes aloquem os *Astyanax* dentro da subfamília Stethaprioninae (Mirande, 2018), essa proposição ainda não possui forte sustentação filogenética e alguns autores consideram o gênero como *Incertae sedis* justamente pela ausência de evidências robustas de monofiletismo (Oliveira *et al.*, 2011; Schmitter-Soto, 2016; Mirande, 2018).

Neste cenário, o sequenciamento de genomas fornece outra via de dados que podem ser utilizados para tentar resgatar as filogenias e se estudar as dinâmicas evolutivas desses organismos. O genoma mitocondrial, em especial, se mostra historicamente promissor quando se tratam de análises filogenéticas (Avise *et al.*, 1987), já que as sequências mitocondriais em si, têm sido muito utilizadas para se estudar até mesmo filogenias mais antigas entre os grupos animais (Curole & Kocher, 1999; Smith, 2015).

Em *Astyanax*, entretanto, abordagens genômicas são bem exploradas apenas em *A. mexicanus*, que é um organismo modelo em estudos evolutivos (Borowski, 2008). Para espécies brasileiras, os estudos genômicos ainda são escassos, de modo que, até o momento, apenas dois mitogenomas foram descritos para o gênero, o de *A. paranae* (Silva *et al.*, 2016) e *A. giton* (Barretos *et al.*, 2017). Assim, neste trabalho, objetivamos descrever o mitogenoma de *Astyanax altiparanae* (uma espécie da bacia do Alto Paraná) e compará-lo com outros genomas mitocondriais disponíveis na literatura.

2.3 Materiais e Métodos:

O *Astyanax altiparanae* utilizado neste trabalho havia sido previamente coletado no Ribeirão Pirapitinga (bacia do rio Paranaíba) em Ibiá – MG (19°21'06.34"S, 46°47'34.81"O) e estava depositado na coleção ictiológica do Laboratório de Genética Ecológica e Evolutiva da Universidade Federal de Viçosa (LaGEEvo), campus Rio Paranaíba-MG, Brasil, sob o voucher 4121 e sob a licença fornecida pelo Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado (SISGEN) A9FE946.

O DNA genômico total foi extraído de amostras de fígado e coração dos indivíduos (obtidas do Banco de Tecidos e Suspensões Celulares do LaGEEvo), conforme as instruções do kit de extração e purificação de DNA PureLink da Invitrogen. Após a verificação da qualidade via NanoDrop Lite Spectrophotometer, o sequenciamento de genoma completo foi realizado em um Novaseq 6000 (Illumina, San Diego, CA) na empresa Novogene, Reino Unido.

A montagem *de novo* com Novoplasty v3.7 foi realizada usando as leituras brutas de 2x150pb do sequenciamento em um computador com clusters paralelos (64 Gb de RAM), com três Kmers distintos (suplemento). As sequências obtidas foram anotadas no MitoAnnotator (Iwasaki *et al.*, 2013) do MitoFish (<http://mitofish.aori.u-tokyo.ac.jp>). Para análises de composição e qualidade do mitogenoma, a sequência obtida foi submetida a um workflow de ferramentas de bioinformática (Figura 1 do material suplementar) disponíveis na plataforma pública do Galaxy europeu (<https://usegalaxy.eu/>, Afgan *et al.*, 2018).

A partir da sequência mitogenômica completa de *A. altiparanae* e de sequências mitocondriais de outras espécies de Characiformes obtidas do GenBank, os 13 genes codificadores de proteínas (PCGs) foram concatenados e utilizados na construção de uma árvore de máxima verossimilhança no software MEGA X (Kumar *et al.*, 2018), com 1.000 réplicas de *bootstrap*, depois de submetidos ao teste de melhor modelo evolutivo (GTR+G+I) também pelo MEGA X (Figura 2).

2.4 Resultados:

O sequenciamento de genoma completo resultou em um conjunto de dados de 33 milhões de *reads* de 150pb de comprimento para o *A. altiparanae* (Tabela 1). Dentro destes dados, o genoma mitocondrial total obtido de *A. altiparanae* contém 16.730pb de comprimento e é composto por 37 genes funcionais (o voucher do GenBank estará disponível após publicação do artigo que está em preparação) (Figura 1), dos quais 13 são genes codificadores de proteínas (PCGs), 22 genes de tRNA e dois genes de rRNA, seguindo a ordem esperada de acordo com os mitogenomas dos outros *Astyanax* já descritos. A região de controle terminal (D-loop) possui 1.067pb. Todos os PCGs estão na cadeia pesada (H), exceto o gene ND6, que está na cadeia leve, juntamente com oito tRNAs (Tabela 2).

O conteúdo de A+T corresponde a 58,3% do mitogenoma. A composição das bases é de 29,59% para A; 28,66% para T; 15,45% para G e 26,28% para C. Além disso, o mitogenoma de *A. altiparanae* apresenta 11 regiões intergênicas, totalizando 92pb. Os 13 PCGs correspondem a 68,25% do genoma mitocondrial de *A. altiparanae* que, mesmo com pequenas alterações quando comparado a outros peixes do gênero, segue a conformação típica dos vertebrados.

Quase todos os PCGs começam com o códon ATG usual e terminam com TAA ou TAG, as exceções se encontram destacadas na Tabela 2. O mitogenoma de *A. altiparanae* apresentou sobreposições que totalizam 26pb em regiões codificantes (entre tRNA-Ile e tRNA-Gln – 2pb; tRNA-Cys e tRNA-Tyr – 1pb; COI e tRNA-Ser – 13pb; ATPase 8 e ATPase 6 – 13pb) (Tabela 2). Foram encontrados 20 polimorfismos de nucleotídeo único (SNPs), e não foram encontrados indels.

Na a árvore de máxima verossimilhança resultante das análises com o MEGA X, considerando como outgroup o *Acestrorhynchus* sp. (Figura 2), nosso *A. altiparanae* foi

recuperado como clado irmão de *A. mexicanus*. Estes dois clados formaram ainda um grupo monofilético com *A. paranae*, com um alto valor de suporte. Entretanto, *Astyanax* como um todo não foi recuperado como monofilético. Quase todos os clados obtiveram altos valores de suporte, exceto o clado (*G. bogotensis* (*A. paranae* (*A. mexicanus* + *A. altiparanae*))).

Tabela 1: Resumo dos dados brutos oriundos do sequenciamento NGS de *A. altiparanae*.

Número de reads sequenciados:	33,665,935
Tamanho médio da sequência:	150 pb
Número total de bases sequenciadas:	5,050 Mb
Conteúdo G+C total:	38,32%

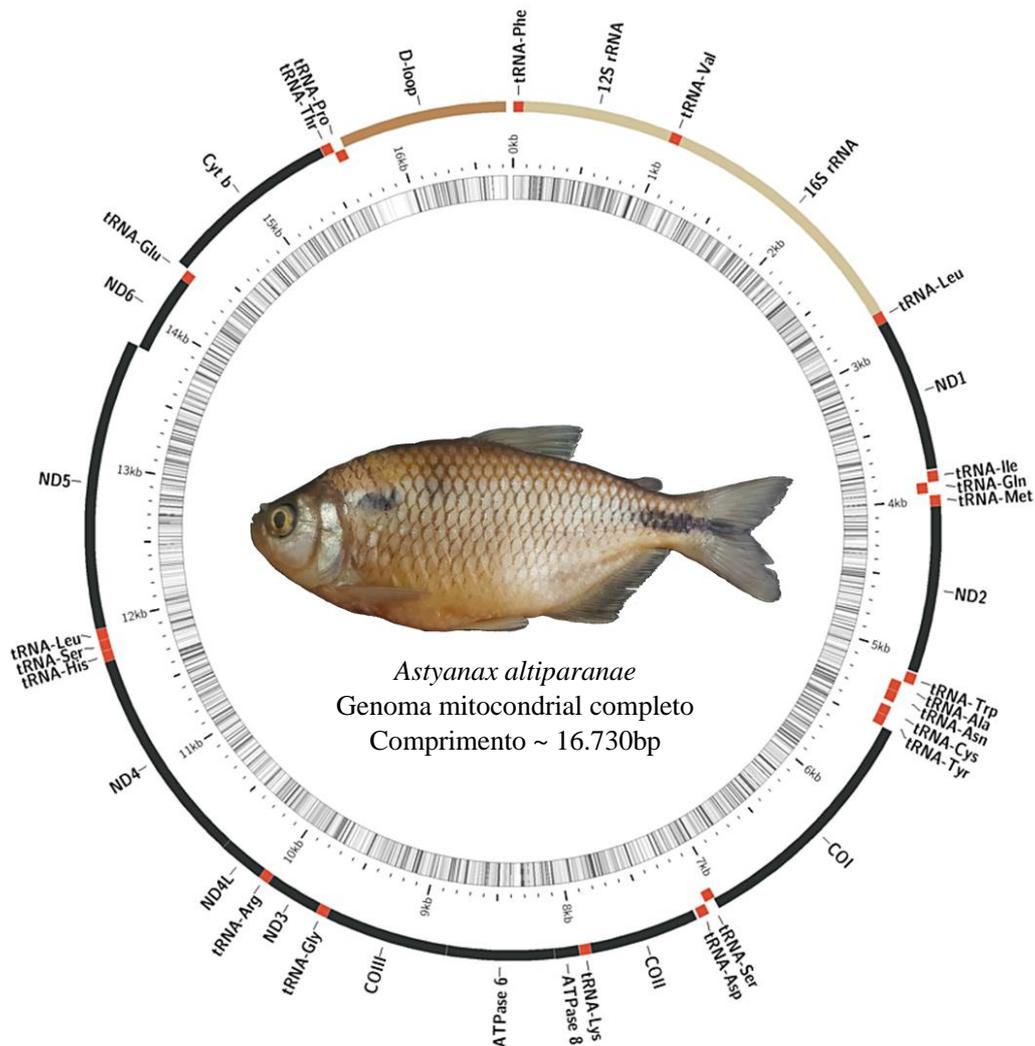


Figura 1: Genoma mitocondrial completo de *A. altiparanae*.

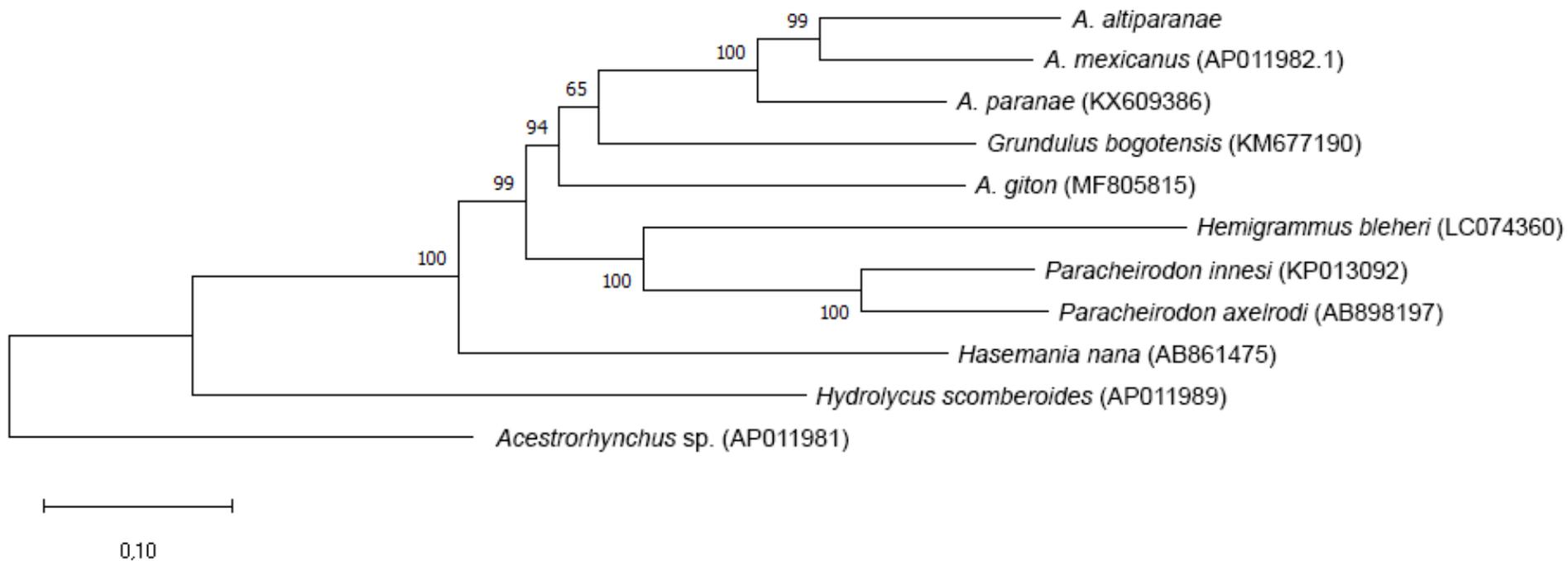


Figura 2: Árvore de máxima verossimilhança obtida através dos 13 PCGs concatenados dos mitogenomas disponíveis de Characidae, sob o modelo GTR+G+I e 1.000 réplicas de Bootstrap. Em parêntese se encontram os vouchers das sequências disponíveis no GenBank.

Tabela 2: Características do genoma mitocondrial de *Astyanax altiparanae*. Os códonos não usuais se encontram grifados e em negrito. As letras H e L indicam, respectivamente, as cadeias pesada (*heavy* - H) e leve (*light* - L).

Gene	Posição		Comprimento (bp)	Cadeia	Códon	
	De	Até			Início	Parada
tRNA ^{Phe}	1	68	68	H		
12S rRNA	69	1014	946	H		
tRNA ^{Val}	1015	1086	72	H		
16S rRNA	1087	2756	1670	H		
tRNA ^{Leu}	2757	2831	75	H		
ND1	2832	3803	972	H	ATG	TAA
tRNA ^{Ile}	3812	3883	72	H		
tRNA ^{Gln}	3882	3953	72	L		
tRNA ^{Met}	3966	3953	70	H		
ND2	3966	4035	1056	H	ATG	TAA
tRNA ^{Trp}	5102	5172	71	H		
tRNA ^{Ala}	5180	5248	69	L		
tRNA ^{Asn}	5250	5322	73	L		
tRNA ^{Cys}	5353	5419	67	L		
tRNA ^{Tyr}	5419	5489	71	L		
COI	5491	7050	1560	H	<u>GTG</u>	<u>AGG</u>
tRNA ^{Ser}	7038	7109	72	L		
tRNA ^{Asp}	7117	7188	72	H		
COII	7203	7893	691	H	ATG	<u>I-</u>
tRNA ^{Lys}	7894	7966	73	H		
ATPase 8	7968	8135	168	H	ATG	TAG
ATPase 6	8126	8807	682	H	<u>GTG</u>	<u>I-</u>
COIII	8808	9591	784	H	ATG	<u>I-</u>
tRNA ^{Gly}	9592	9662	71	H		
ND3	9663	10011	349	H	ATG	<u>I-</u>
tRNA ^{Arg}	10012	10080	69	H		
ND4L	10081	10377	297	H	ATG	TAA
ND4	10371	11751	1381	H	ATG	<u>I-</u>
tRNA ^{His}	11752	11820	69	H		
tRNA ^{Ser}	11821	11888	68	H		
tRNA ^{Leu}	11890	11962	73	H		
ND5	11963	11801	1839	H	ATG	TAA
ND6	13798	14313	516	L	ATG	TAG
tRNA ^{Glu}	14314	14381	68	L		
Cyt b	14387	15523	1137	H	ATG	TAA
tRNA ^{Thr}	15525	15596	72	H		
tRNA ^{Pro}	15595	15663	69	L		
D-loop	15664	16730	1067	H		

2.5 Discussão e Conclusões:

Filogenias precisas são a base para uma compreensão cada vez mais completa da evolução dos vertebrados (Curole *et al.*, 1999). Tanto táxons irmãos, quanto membros de complexos de espécies que possuem tempos de divergência curtos entre si, e que retêm um alto grau de similaridade morfológica, podem se mostrar como espécies crípticas de difícil reconhecimento, mesmo que estejam em diferentes trajetórias evolutivas (Struck *et al.*, 2018). Isso interfere diretamente na construção das filogenias, onde a taxonomia não só desempenha um papel fundamental em uma identificação mais precisa, como também na conservação dos organismos, afinal, não podemos esperar conservar organismos que não podemos nem mesmo identificar (Mace, 2004).

A questão biológica apresentada acima é uma das principais polêmicas levantadas quando se trata dos *Astyanax*, onde nem a tribo (Grundulini - Fowler, 1958) e nem a subfamília (Stethaprioninae) propostas para o gênero, possuem medidas de suporte altas nas análises filogenéticas moleculares e morfológicas (Mirande, 2018). Entretanto, estudos anteriores mostraram que populações alopátricas de *Astyanax*, como em *A. scabripinnis* e *A. bimaculatus* podem mostrar estruturação cromossômica associada ou não à estruturação morfológica, onde é possível diferenciar suas populações (Pazza *et al.*, 2008b & 2016; Fernandes *et al.*, 2014). Neste cenário, combinações de diferentes tipos de caracteres taxonômicos são necessárias para uma taxonomia mais integrativa e resolutiva e, conseqüentemente, filogenias mais robustas (De Queiroz, 2007; Padial *et al.*, 2010; Struck *et al.*, 2018).

Na filogenia mitocondrial proposta por Silva e colaboradores em 2016, a espécie mais próxima de *A. paranae* era *A. mexicanus*. A topologia dessa árvore, com a inserção do nosso genoma mitocondrial de *A. altiparanae*, se torna ligeiramente diferente. Em nossas análises (Figura 2), *A. altiparanae* passa a ser o grupo irmão mais próximo de *A. mexicanus*, com *A. paranae* como irmão do clado por eles formado. Essa nova topologia é concordante com a filogenia baseada na sequência da ATPase6/8, em associação a marcadores cromossômicos, proposta por Pazza *et al.* (2018). Comparando com outros characiformes, *Astyanax* continua como parafilético, como foi observado por Barreto *et al.* (2017), em função das relações evolutivas de *A. giton*.

De qualquer forma, a organização final dos genes mitocondriais obtidos de *A. altiparanae* é semelhante à dos outros *Astyanax* disponíveis na literatura, com diferenças no comprimento e na composição de alguns genes. O D-loop de *A. altiparanae*, por exemplo, tem 70pb a mais que *A. giton* e 35pb a mais que *A. paranae*.

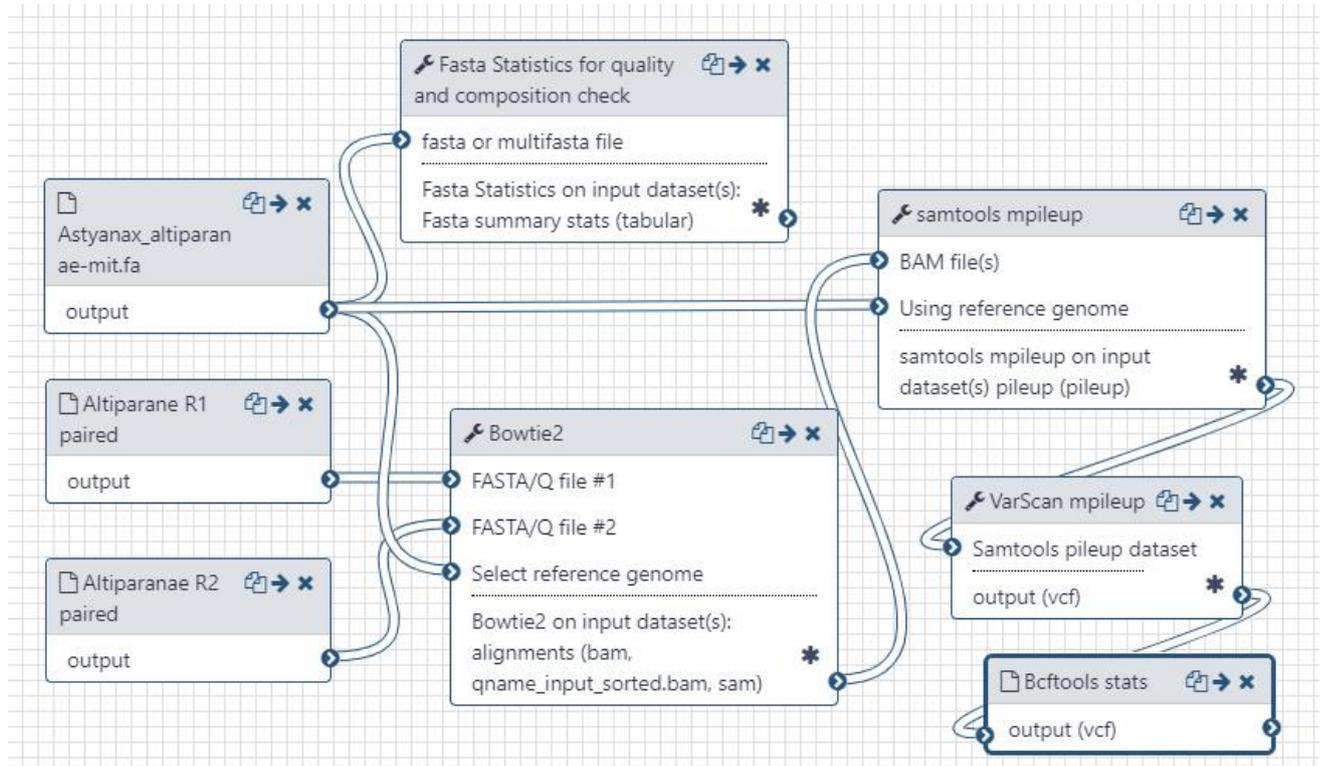
O mitogenoma de *A. altiparanae* é o maior em comprimento dentre os *Astyanax* da literatura, ainda assim, seus PCGs são ligeiramente menores que o de *A. giton* (3,55%). O tamanho maior do mtDNA de *A. altiparanae* provavelmente se deve ao longo D-loop e às regiões intergênicas, que totalizam 92pb, quantidade maior que a encontrada em outras espécies da família Characidae (Barreto *et al.*, 2017).

Incongruências filogenéticas ainda são um problema mesmo na era dos grandes dados genômicos (Betancur *et al.*, 2019). Como as sequências mitocondriais oferecem naturalmente uma série de vantagens sobre outros genes nas análises filogenéticas (Avice *et al.*, 1987), o uso de genomas mitocondriais completos tem resultado na recuperação de filogenias bem resolvidas em diversos animais (de Freitas *et al.*, 2018; Fu *et al.*, 2019).

Embora a maioria dos mtDNAs atualmente sequenciados sejam de animais (Smith, 2015), os *Astyanax* ainda carecem desse tipo de marcador genético em suas filogenias. Deste modo, concluímos que o mitogenoma completo de *A. altiparanae* disponibilizado, poderá não só auxiliar numa taxonomia mais precisa dentro do complexo *A. bimaculatus*, como também em estudos futuros de genética populacional e contribuirá para a resolução da complexa filogenia do gênero, onde nossos resultados apoiam a condição parafilética para os *Astyanax*, por enquanto, pois novas hipóteses poderão ser testadas à medida que mais mitogenomas forem incorporados nas análises.

2.6 Material Suplementar:

Figura 1: Workflow gerado na plataforma Galaxy mostrando as etapas e ferramentas utilizadas. Para análise de qualidade do genoma mitocondrial de *A. altiparanae* foi utilizado o Fasta Statistics. A busca por SNPs e InDels foi performada mapeando os reads R1 e R2 contra a sequência mitocondrial através das ferramentas Bowtie2, Samtools mpileup, VarScan mpileup e o arquivo final foi conferido e visualizado pelo Bcftools stats.



Referencial teórico das ferramentas de bioinformática:

Afgan E, Baker D, Van Den Beek M, Blankenberg D, Bouvier D, Čech M, ... & Gruning B: The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research*, 2016; 44-W1, W3-W10

Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Čech M, ... & Guerler A: The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research*, 2018; 46-W1, W537-W544.

Danecek P, Schiffels S, Durbin R: Multiallelic calling model in bcftools (-m), 2016. [\[Link\]](#)

Durbin R: Segregation based metric for variant call QC, 2014. [\[Link\]](#)

Koboldt D C, Zhang Q, Larson D E, Shen D, McLellan M D, Lin L, Miller C A, Mardis E R, Ding L, Wilson R K: VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. In *Genome Research*, 2012; 22-3, pp. 568-576. [\[Link\]](#)

Koboldt D C, Larson D E, Wilson R K: Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection. In Current Protocols in Bioinformatics, 2013; pp. 15.4.1â€“15.4.17. [[Link](#)]

Langmead B, Trapnell C, Pop M, Salzberg S L: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. In Genome Biology, 2009; 10-3, pp. R25. [[Link](#)]

Langmead B, Salzberg S L: Fast gapped-read alignment with Bowtie 2. In Nature Methods, 2012; 9-4, pp. 357â€“359 [[Link](#)]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: The Sequence Alignment/Map format and SAMtools. In Bioinformatics, 2009; 25-16, pp. 2078â€“2079. [[Link](#)]

Li H: Improving SNP discovery by base alignment quality. In Bioinformatics, 2011; 27-8, pp. 1157â€“1158 [[Link](#)]

Li H: A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. In Bioinformatics, 2011; 27-21, pp. 2987â€“2993 [[Link](#)]

Li H: Mathematical Notes on SAMtools Algorithms, 2010 [[Link](#)]

SAMTools GitHub page [[Link](#)]

Torsten S & Simon Gladman: Fasta Statistics: Display summary statistics for a fasta file, 2012 [[Link](#)]

3. CAPÍTULO 2:

IDENTIFICAÇÃO *DE NOVO* E INVESTIGAÇÃO *IN SILICO* DO SATELITOMA DE *Astyanax altiparanae* (CHARACIFORMES, CHARACIDAE)

Matheus Lewi Bonaccorsi de Campos^{1,2,5}, Fabiano Bezerra Menegidio³, John Seymour (Pat) Heslop-Harrison⁴, Trude Schwarzacher⁴, Rubens Pasa^{2,5}, Karine Frehner Kavalco^{2,5}

¹ Programa de Pós-Graduação em Zoologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.

² Laboratório de Genética Ecológica e Evolutiva, Universidade Federal de Viçosa, campus Rio Paranaíba, Rio Paranaíba, MG, Brasil.

³ Núcleo Integrado de Biotecnologia, Universidade de Mogi das Cruzes, Brasil.

⁴ Department of Genetics and Genome Biology, University of Leicester, United Kingdom.

⁵ Laboratório de Bioinformática e Genômica, Universidade Federal de Viçosa, campus Rio Paranaíba, Brasil.

3.1. Resumo:

DNAs satélites são sequências repetidas em tandem formando matrizes que podem ir de dois pares de bases até megabases de comprimento. Eles estão entre os elementos repetitivos mais abundantes e de evolução mais rápida do genoma eucarioto, onde podem desempenhar uma série de papéis funcionais nos cromossomos. O barateamento das tecnologias de sequenciamento, em conjunto com o desenvolvimento recente de softwares específicos para estudos de satelitoma (conjunto de satDNAs em um genoma), têm impulsionado o uso de sequências repetitivas como marcadores citogenéticos, os quais ajudam no entendimento sobre como essas sequências podem influenciar a organização e a evolução dos genomas e das espécies, mesmo em organismos não-modelo. Neste trabalho, utilizamos dos softwares de agrupamento RepeatExplorer e TAREAN, auxiliados pelos bancos de dados do Blast e do Dfan, para realizar a identificação *de novo* e o estudo *in silico* de sequências repetitivas do genoma de *Astyanax altiparanae* (obtido via NGS), em busca de possíveis marcadores citogenéticos. Encontramos um total de 32 satélites putativos, dos quais 15 são correspondentes a sequências já descritas e 17 são sequências inteiramente novas que requerem validação via FISH, e concluímos que pelo menos 3 dos satélites putativos são marcadores citogenéticos em potencial.

3.2.Introdução:

Os DNAs satélites (satDNAs) estão entre os elementos repetitivos mais abundantes do genoma eucarioto (Silva *et al.*, 2019). SatDNAs são sequências repetidas em tandem, formando matrizes que podem ir de dois pares de bases até megabases de comprimento (López-Flores & Garrido-Ramos, 2012; Silva *et al.*, 2019). Eles podem ser classificados em microssatélites (até 9 nucleotídeos encontrados em um kilobase), minissatélites (mais de 9 nucleotídeos em 1kbp) e satélites (200 nucleotídeos ou mais) à depender de seu comprimento, mas não há um consenso preciso sobre esses limites de tamanho (Tautz, 1993; Richard *et al.*, 2008; Ruiz-Ruano *et al.*, 2016).

SatDNAs se encontram distribuídos preferencialmente em regiões heterocromáticas pericentrométricas ou subteloméricas dos cromossomos, podendo também serem encontrados dispersos na eucromatina (Kuhn *et al.*, 2012; Pavlek *et al.*, 2015; Utsunomia *et al.*, 2019). Apesar de geralmente não serem sequências codificantes, ainda assim, satDNAs podem desempenhar uma série de papéis funcionais nos cromossomos, relacionados à estruturação cromossômica, modulação da cromatina, estabelecimento de centrômeros ou mesmo regulação gênica (Rosic *et al.*, 2014; Bracewell *et al.*, 2019; Silva *et al.*, 2019).

O conjunto de famílias de DNAs satélites pode ser chamado de Satelitoma (Ruiz-Ruano *et al.*, 2016), sendo um dos grupos de DNAs repetitivos de evolução mais rápida do genoma (Kuhn *et al.*, 2008; Ruiz-Ruano *et al.*, 2016; Duílio *et al.*, 2017). Alguns componentes do satelitoma de um organismo podem ser específicos de um cromossomo ou de uma espécie, e cada espécie apresenta uma biblioteca inteira de satDNAs que pode ou não ser compartilhada com táxons mais próximos (Vicari *et al.*, 2010; Rodrigues *et al.*, 2019).

Desde a descoberta dos primeiros DNAs satélites nos anos 60, diversos estudos sobre a organização, composição e localização física desses DNAs repetitivos nos cromossomos elucidaram várias questões sobre mecanismos evolutivos do genoma: as altas taxas evolutivas do satelitoma, sua abundância e papel estrutural possuem todas implicações importantes para a diversificação e evolução dos genomas e das espécies (Kursel *et al.*, 2018; Silva *et al.*, 2019). Considerando que organismos proximamente relacionados podem compartilhar várias famílias de satDNAs (Vicari *et al.*, 2010), que geralmente evoluem independentemente dentro das diferentes linhagens, esses elementos

repetitivos fornecem excelentes meios para estudos evolutivos e filogenéticos (Vicari *et al.*, 2010; Utsunomia *et al.*, 2019).

Apesar dos crescentes estudos sobre DNAs satélites, ainda se sabe pouco sobre a totalidade de suas funções quando comparados aos estudos que envolvem outros repetitivos, como por exemplo o DNA ribossômico (Mehrotra *et al.*, 2014; Pamponét *et al.*, 2019). Isso talvez se deva ao histórico dos complexos métodos laboratoriais mais tradicionais de obtenção desse tipo de DNA em inúmeros organismos (Telenius *et al.*, 1992; Heslop-Harrison, 2000; Ferreira & Martins *et al.*, 2008; Vicari *et al.*, 2010).

Nos últimos anos, com o avanço dos recursos computacionais, houve uma explosão de softwares mais acessíveis e dedicados a melhorar nossa capacidade de reconhecer, identificar, comparar e avaliar sequências repetitivas em genomas inteiros (Lower *et al.*, 2018). Entretanto, ainda assim, os satDNAs têm sido pouco estudados em organismos não-modelo (Treangen & Salzberg, 2011), para os quais as tecnologias de sequenciamento mais baratas e acessíveis apresentam uma série de limitações. Algumas delas, como a complexidade das técnicas de bioinformática (principalmente para profissionais que não possuem formação na área) (Afgan *et al.*, 2018) e a alta similaridade e a variação no número de cópias que as sequências repetitivas apresentam, tornam difíceis a identificação computacional precisa, a montagem do genoma e a anotação funcional dessas sequências altamente repetitivas como os satDNAs (Treangen & Salzberg, 2011; Afgan *et al.*, 2018; Lower *et al.*, 2018).

As abordagens que utilizam do sequenciamento *shotgun* de genoma completo, porém, têm avançado e sido cada vez mais utilizadas para contornar os problemas citados acima (Lower *et al.*, 2018). O sequenciamento de genoma completo de nova geração (NGS) em *reads* (leituras) de baixa cobertura, não só ajuda a evitar os possíveis vieses de montagem genômica (Lower *et al.*, 2018), como também permite a identificação imparcial de milhares de repetições, tornando possível o estudo *in silico* do satelitoma (Mascagni *et al.*, 2020), através de softwares de agrupamento como o RepeatExplorer e o Tandem Repeat Analyzer (TAREAN) (Nóvak *et al.*, 2010; 2013; 2017).

Esses softwares rodam *pipelines* computacionais que permitem a identificação *de novo* (do latim: do início) não supervisionada de sequências repetitivas a partir de leituras não montadas. A *pipeline* usa as sequências inteiras do genoma para executar uma ordem lógica de operações: 1- Identificação e agrupamento, com base em gráficos de Bruijn, de

leituras sobrepostas que são derivadas de sequências repetitivas; 2- A partir dos gráficos gerados, é feita uma atribuição inicial dessas leituras a famílias de repetitivos; 3- Quantificação da abundância genômica e da variação das sequências de cada repetitivo, sendo geradas sequências consenso; 4- Classificação final dos repetitivos com base nos monômeros repetidos, nos tamanhos das repetições de cada cluster agrupado e no layout final dos gráficos em relação a um banco de dados do sistema; 5- Criação de um relatório interativo em HTML com todos os dados gerados (Heslop-Harrison, 2000; Nývák *et al.*, 2010; 2013; 2017; Mehrotra & Goyal, 2014; Lower *et al.*, 2018).

Avaliar, quantificar e identificar com precisão os componentes do satelitoma de todo um genoma em diferentes táxons, requer o desenvolvimento de abordagens alternativas (Lower *et al.*, 2018) e como a ictiofauna neotropical é extremamente rica e diversa, são necessários testes e o subsequente uso combinado de diferentes métodos quando se tratam de estudos sobre composição, diversificação e evolução dos genomas (Vicari *et al.*, 2010;).

Abordagens *in silico* têm se mostrado úteis e eficazes na busca por marcadores citogenéticos tanto em plantas (Pamponét *et al.*, 2019; Mascagni *et al.*, 2020), quanto em animais como gafanhotos (cujo genoma é um dos maiores em pb) (Ferretti *et al.*, 2019), moscas (Silva *et al.*, 2019) e peixes (Ruiz-Ruano *et al.*, 2016; Duílio *et al.*, 2017). Estudos anteriores com satDNAs em peixes também já demonstraram o grande potencial que esse tipo de sequência tem para a investigação sobre diferenciação cariotípica e evolução da ictiofauna neotropical (Vicari *et al.*, 2010; Barbosa *et al.*, 2017).

Em peixes Teleósteos, por exemplo, satDNAs específicos de cromossomos, de espécies ou de gêneros, como o As-51 dos *Astyanax* (Mestriner *et al.*, 2000), são marcadores citogenéticos extremamente úteis que ajudam à identificar e rastrear rearranjos cromossômicos e conseqüentemente à traçar suas filogenias (Mestriner *et al.*, 2000; Saito *et al.*, 2007; Pazzo *et al.*, 2018). Sendo assim, objetivamos com este trabalho realizar a prospecção de marcadores citogenéticos para o gênero *Astyanax*, investigando *in silico* o satelitoma gerado por NGS a partir de amostras de um exemplar de *A. altiparanae*.

3.3. Materiais e métodos:

O *Astyanax altiparanae* utilizado neste trabalho havia sido previamente coletado no Ribeirão Pirapitinga (bacia do rio Paranaíba) em Ibiá – MG (19°21'06.34"S, 46°47'34.81"O) e estava depositado na coleção ictiológica do Laboratório de Genética Ecológica e Evolutiva da Universidade Federal de Viçosa (LaGEEvo), campus Rio Paranaíba-MG, Brasil, sob o voucher 4121 e sob a licença fornecida pelo Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado (SISGEN) A9FE946.

O DNA genômico total foi extraído de amostras de fígado e coração dos indivíduos (obtidas do Banco de Tecidos e Suspensões Celulares do LaGEEvo), conforme as instruções do kit de extração e purificação de DNA PureLink da Invitrogen. Após a verificação da qualidade via NanoDrop Lite Spectrophotometer, o sequenciamento de genoma completo foi realizado em um Novaseq 6000 (Illumina, San Diego, CA) na empresa Novogene, Reino Unido.

As leituras brutas de 2x150pb foram carregadas para os servidores europeus da plataforma Galaxy (<http://usegalaxy.eu/> - Afgan *et al.*, 2018), onde foram utilizadas de sucessivas ferramentas de bioinformática para checagem e análise dos dados. A qualidade do genoma foi conferida através do FastQC Read Quality Reports (Andrews, 2010) e foi utilizado o Trimmomatic (Bolger *et al.*, 2014) para remoção das sequências adaptadoras e contaminantes.

Para a identificação *in silico* dos possíveis DNAs satélite, as leituras pareadas resultantes do Trimmomatic foram submetidas às *pipelines* do RepeatExplorer (<https://repeatexplorer-elixir.cerit-sc.cz/> - Novak *et al.*, 2013-2) e do Tandem Repeat Analyzer (TAREAN - Novak *et al.*, 2017), implementados na plataforma Galaxy (Figura 1 do material suplementar). Através da ferramenta Preprocessing of fastq paired reads (incluída na instância RepeatExplorer do Galaxy) foram excluídas as leituras únicas e as que continham mais de 5% de seus pares de bases em baixa qualidade (Tabela 1 do material suplementar).

Ainda com a Preprocessing, as leituras foram emparelhadas e entrelaçadas, conforme os requisitos necessários para uso do RepeatExplorer. O arquivo único gerado pelo pré-processamento foi então utilizado como arquivo de entrada para o agrupamento

por similaridade realizado pelo RepeatExplorer e pelo TAREAN (Tabela 2 do material suplementar).

A separação das leituras em diferentes clusters, que correspondem à diferentes famílias de elementos repetitivos, é feita automaticamente pela *pipeline* do RepeatExplorer. Ele realiza a decomposição das leituras em k-mers, cuja frequência de repetição e os níveis de similaridade são utilizados para compor os vértices e arestas dos gráficos de Bruijn. A disposição desses gráficos é ordenada conforme as sobreposições e alinhamentos necessários para gerar as sequências consenso, onde o layout final do gráfico é analisado e classificado estatisticamente pelo TAREAN.

Os clusters com repetições em tandem identificadas então pelos algoritmos do TAREAN são classificados automaticamente em quatro principais categorias: supostos satélites de alta confiança; supostos satélites de baixa confiança; potenciais elementos LTR (Repetições Terminais Longas das extremidades de retrotransposóns); e rDNAs.

As estimativas de alta ou baixa confiança dos possíveis satélites são atribuídas pelo TAREAN de acordo com o Índice de Componentes Conectados (C) e com o Índice de Completude dos Pares (P) presentes no programa. Esses índices são utilizados simultaneamente para detectar e separar as repetições putativas de DNAs satélites, onde os valores ideais de confiabilidade esperados para ambos são próximos de 1.

As sequências consenso de cada um dos clusters de possíveis DNAs satélites, identificadas pelo combo RepeatExplorer e TAREAN, foram dispostas no banco de dados do Blast (Altschul *et al.*, 1990, disponível em <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) e do Dfam 3.1 (Hubley *et al.*, 2016, disponível em <https://dfam.org/home>), para busca de similaridades e possíveis homologias com outras sequências já descobertas.

Para o Blast foi utilizada a otimização do algoritmo Megablast e no Dfam foi utilizada a espécie *Danio rerio* (Peixe-Zebra) como organismo modelo para Actinopterygii. Tanto no Dfam quanto no Blast foram considerados bons alinhamentos aqueles com E-value (probabilidade devido ao acaso de que exista outro alinhamento com uma similaridade maior que a medida de semelhança da pesquisa com a sequência mostrada) de e^{-5} ou inferior, com uma cobertura acima de 70%.

Para testes comparativos, a sequência do satélite AalSat24-51 foi utilizada no software MEGA X, onde foram feitos alinhamentos individuais (algoritmo Muscle) dessa

sequência com a de outros satélites já descritos para *Astyanax*, disponíveis no GenBank: ApaSat01-51 (MF044760.1), ApaSat15-51 (MF044784.1) e satDNA As-51 (U87962.1).

3.4. Resultados e discussões:

Estudos com DNAs satélites têm sido realizados em *Astyanax* há décadas, ainda assim, o As-51 identificado em *A. scabripinnis* era o único satélite descrito para o gênero desde os anos 2000 (Mestriner *et al.*, 2000). Era, pois, recentemente, com dados de NGS e softwares específicos para a avaliação de satDNAs, como o próprio RepeatExplorer, outros 45 satélites já foram descritos para *A. paranae* e testados em *A. fasciatus* e *A. bockmanni* (Duílio, *et al.*, 2017).

Em nossas análises, a saída do RepeatExplorer com o TAREAN foi fornecida na forma de um arquivo em HTML (*HyperText Markup Language*, documentos em HTML podem ser abertos diretamente em navegadores web), que inclui um relatório e uma lista de todos os clusters analisados. Os nomes dos satélites então identificados receberam a seguinte estrutura: Aal (abreviação do prefixo referente à espécie *Astyanax altiparanae*) + Sat + número de identificação do cluster em ordem decrescente de abundância + comprimento do monômero consenso em pares de bases (Ruiz-Ruano *et al.*, 2016).

No presente trabalho, O sequenciamento de genoma completo resultou em um conjunto de dados de 33 milhões de *reads* de 150pb de comprimento para o *A. altiparanae* (Tabela 1). Em nossa investigação *in silico* do satelitoma de *Astyanax altiparanae* via RepeatExplorer e TAREAN resultou na descoberta de 32 satélites putativos (Tabela 2), com um total de 11.061 pares de bases, representando juntos aproximadamente 3,5% do genoma de *A. altiparanae* (o voucher do GenBank estará disponível após publicação do artigo). Não foram encontrados elementos LTR e nem sequências repetitivas de rDNA nessa análise.

Como os satDNAs estão entre os elementos repetitivos mais abundantes de um genoma (Charlesworth *et al.*, 1994), representando uma parcela significativamente alta do genoma em diversos animais (López-Flores & Garrido-Ramos, 2012), a porcentagem real de sequências repetitivas no genoma de *A. altiparanae* deve ser maior do que os 3,5% aqui encontrados.

Esse valor provavelmente subestimado se dá em função dos algoritmos de agrupamento por similaridade. Além de não detectarem repetitivos de baixa abundância (Novák *et al.*, 2017), eles podem acabar classificando monômeros curtos e de sequências

parecidas, mas que são fundamentalmente diferentes, em um mesmo cluster, como um mesmo tipo de elemento repetitivo (Lower *et al.*, 2018; Silva *et al.*, 2019). Nesse sentido, o TAREAN sozinho não é o melhor método para uma quantificação generalizada, principalmente de satélites com repetições curtas (Novák *et al.*, 2017). Em *Drosophila*, por exemplo, alguns autores estimaram, a partir de métodos de ultracentrifugação em gradiente de densidade, que até 40% do genoma seja de satDNAs (Gall *et al.*, 1971; Cohen *et al.*, 1979), ao passo que Silva e colaboradores (2019) encontraram um máximo de 12% desses repetitivos no gênero utilizando do TAREAN.

Apesar disso, o RepeatExplorer e o TAREAN têm sido amplamente utilizados para estudos de DNAs repetitivos, principalmente em plantas e organismos não-modelo (Ruiz-Ruano *et al.*, 2016; Pamponét *et al.*, 2019), pois não só não requerem o uso de um genoma de referência (Novák *et al.*, 2017), como também fornecem resultados rápidos e detalhados em uma interface interativa relativamente simples, dentro de um servidor web gratuito com alta capacidade de processamento e armazenamento (Goecks *et al.*, 2010). Ressalto que armazenamento e capacidade de processamento são um dos maiores limitantes para se trabalhar com *Genomic Data Science*, onde um sequenciador pode gerar terabytes de dados em poucas horas (Goecks *et al.*, 2010; Afgan *et al.*, 2016).

As sequências dos nossos satDNAs putativos, em geral, variaram de 19 a 3.220 pb de comprimento. Dessas, 6 sequências de supostos satélites foram classificadas como sendo de alta confiança e as demais 26 como de baixa confiança. É interessante ressaltar, porém, que esses parâmetros nem sempre se referem estritamente à confiabilidade em si de uma sequência ser ou não um satélite.

Como discutido por Novák *et al.* (2017), o índice C permite a identificação de aglomerados gráficos que correspondem a sequências repetidas em tandem, porém, ele não discrimina com eficiência DNAs satélites dos outros tipos de repetições em tandem, onde entraria o índice P. A partir do momento em que o índice P mede a proporção entre os pares de *reads* completos em um cluster e o número de *reads* quebrados, em relação ao comprimento total das repetições em tandem contínuas, ambos os índices em conjunto, ou seja, a Probabilidade do Satélite, também pode indicar se os repetitivos se dispõem agrupados em tandem ou se estão dispersos no genoma (Silva *et al.*, 2019).

Uma das características computacionais de repetições do tipo satélite para os índices C e P em questão, é que elas contêm proporções baixas de pares de *reads*

quebrados, ao passo que as proporções de *reads* quebrados são maiores nos repetitivos que estão espalhados pelo genoma (Novák *et al.*, 2017). Neste caso, os 26 SatDNAs com valores de Probabilidade do Satélite menores que 0,7, que são então denotados como sendo de baixa confiança pelo TAREAN, podem na verdade ser sequências de repetitivos que se encontram dispersos no genoma ou que possuem características intermediárias entre os diferentes tipos de repetitivos (Novák *et al.*, 2017; Lower *et al.*, 2018; Silva *et al.*, 2019).

3.4.1. Classificação dos satélites putativos em relação ao banco de dados do Blast e NCBI:

Dos 32 satélites putativos encontrados no satelitoma de *A. altiparanae*, dezessete não apresentaram correspondência com nenhum banco de dados, sendo inteiramente novos. Uma vez que essas sequências carecem de quaisquer outras informações na literatura, elas precisariam ser estudadas além das abordagens *in silico*, a partir das técnicas de FISH, para sua caracterização, validação e melhor entendimento. Os outros quinze satélites putativos apresentaram algum grau de correspondência com outras sequências já disponíveis nos bancos de dados do Blast e do Dfam (Tabela 2). No Blast, 7 apresentam similaridade com outros 7 DNAs satélites de *Astyanax paranae* e 2 corresponderam a diferentes tipos de mRNA de *A. mexicanus*.

O nosso satélite putativo AalSat01-178 apresentou uma abundância de 0,54% sendo então o mais abundante no genoma de *A. altiparanae*. Dados os altos valores de confiança e probabilidade deste satélite, ele provavelmente está arranando em um cluster, hipótese esta que pode ser testada futuramente através da FISH. Comparativamente, sua sequência apresentou 95% de identidade com a do satélite ApaSat10-179, que seria o décimo em abundância no genoma do *A. paranae* (Duílio, *et al.*, 2017). Os satDNAs putativos AalSat20-22 e AalSat31-27 (um dos satélites de menor abundância em *A. altiparanae*) apresentaram 100% de identidade com o ApaSat11-22 e ApaSat25-27 respectivamente, o alto E-value desse comparativo provavelmente se deve ao diminuto tamanho de suas sequências. Infortunadamente, para mais análises comparativas, os dois últimos satélites, em conjunto com o ApaSat10-179, ainda carecem de mais informações na literatura (Duílio, *et al.*, 2017).

Em contrapartida, com relação ao putativo AalSat11-236, ele obteve 86% de correspondência com o ApaSat04-223. Este último satDNA já foi encontrado formando clusters na região pericentromérica de todos os cromossomos de *A. fasciatus* (incluindo cromossomos B), onde a superfamília deste satélite, a SF1, também compõe a mesma região na maioria dos cromossomos do complemento A e B de *A. paranae* e *A. bockmanni* (com marcações mais sutis nesta última) (Duílio, *et al.*, 2017). Sendo assim, apesar de ser bem menos abundante no genoma de *A. altiparanae* do que o ApaSat04-233 é no genoma dos outros *Astyanax*, com esse nível de similaridade entre os dois satélites, é possível que o AalSat11-236 também desempenhe algum papel de organização centromérica nos cromossomos de *A. altiparanae* (Metters *et al.*, 2013; Ruiz-Ruano *et al.*, 2016).

O satélite AalSat14-229 obteve diferentes correspondências nos bancos de dados. Enquanto o TAREAN o classificou como sendo 0,72% similar à um rDNA 5S, o Blast mostrou 98% de correspondência dessa sequência com uma sequência de rDNA 5S + NTS (espaçador não transcrito) de *A. altiparanae*, ao passo que o Dfam classificou esse satélite como pseudogene de rDNA. Normalmente, o rDNA 5S consiste em sequências codificantes de 120pb de comprimento repetidas em tandem, sendo separadas uma da outra por uma sequência espaçadora não transcrita (NTS) de comprimento variável (Martins & Galetti, 2001). Diferentes classes de rDNA 5S já foram reportados para espécies de peixes, sendo alguns de seus variantes relacionados à pseudogenes e à diferenças na sequência NTS.

A sequência consenso obtida para o AalSat14-229, conforme classificado pelo Blast, tinha uma composição de bases similar à do rDNA 5S descrito para *A. altiparanae* (Hashimoto *et al.*, 2011), entretanto, como tanto a ordem dos genes codificantes quanto da provável sequência NTS eram significativamente diferentes, nós o consideramos como um pseudogene, conforme classificado pelo Dfam. Sendo um pseudogene, pode ser mais útil utilizar o rDNA 5S ao invés do putativo AalSat14-299, de qualquer forma, tal qual o próprio rDNA 5S, cujas altas variações do NTS têm sido bastante úteis em estudos evolutivos (Martins & Galetti, 2001; Pazza *et al.*, 2018), o AalSat14-229 pode vir à ser um candidato à marcador citogenético (caso sua presença seja comprovada via de FISH) e ou molecular, onde pode ser útil na citotaxonomia, ajudando à caracterizar populações de *A. altiparanae*.

3.4.2.O putativo AalSat24-51 como um possível variante do As-51:

Outro satélite que requer distinção é o putativo AalSat24-51. Pelo Blast, sua sequência apresentou 100% de identidade e 75% de cobertura com o ApaSat15-51, sendo o conteúdo AT de ambos de 54,90%. O comprimento de 51pb destes satélites em conjunto com seu conteúdo AT rico se destacam, pois, estas são características similares às do DNA satélite mais estudado e utilizado como marcador citogenético no gênero *Astyanax*, o As-51 (Mestriner *et al.*, 2000; Kantek *et al.*, 2009; Vicari *et al.*, 2010; Pazza *et al.*, 2018).

Historicamente, o As-51 apresenta uma variedade considerável em número e posição nos cromossomos de diferentes espécies e populações de *Astyanax* (Vicari *et al.*, 2010). Alguns autores discutem que como o As-51 provavelmente se originou de elementos transponíveis (Mestriner *et al.*, 2000), mecanismos intrínsecos de amplificação da sua sequência e de transposição para outros locais cromossômicos, permitiram que diferentes populações ou espécies demonstrem diversos padrões de hibridação desse repetitivo (Pazza *et al.*, 2018), indicando também que seguiram por caminhos evolutivos distintos (Kantek *et al.*, 2009; Vicari *et al.*, 2010; Pazza *et al.*, 2018).

Em *A. mexicanus*, foram encontradas marcações muito sutis do As-51 em poucos locais cromossômicos, de modo mais difuso do que nos outros *Astyanax*, isso sugere um número pequeno de cópias do satélite ou mesmo homologia parcial com a sonda (Kavalco e Almeida-Toledo 2007). Já nas espécies dos complexos *A. scabripinnis* e *A. fasciatus*, esse satélite tem sido bem mais evidenciado, com sua presença variando de 0 a 18 sítios marcados (Vicari *et al.*, 2010; Kavalco *et al.*, 2013; Pazza *et al.*, 2018). Para os *A. fasciatus*, estudos prévios demonstraram que as populações das drenagens dos rios Grande, Tietê e Paranapanema apresentaram um maior número de marcações para o As-51 (de 10 à 18 cromossomos marcados), sendo este satélite mais disperso nos genomas, ao passo que as populações de bacias mais costeiras ao leste do continente, apresentaram um número significativamente menor de marcações (de 2 à 5 cromossomos marcados) (Kavalco *et al.*, 2013).

Essas diferenças nas populações de *A. fasciatus* estão relacionadas à um padrão biogeográfico em que quanto maior a distância dos limites da bacia do Alto Paraná e mais próximo das bacias hidrográficas costeiras, menor é o número de sítios do As-51 (Kantek *et al.*, 2009; Peres *et al.*, 2009; Kavalco *et al.*, 2013). É interessante destacar que essas populações de *A. fasciatus*, caracterizadas por poucas marcações do As-51, apresentam

sua distribuição ao longo da costa brasileira em conjunto com outras espécies de bacias costeiras que não demonstram marcação para o satélite como *A. intermedius*, *A. giton* e *A. hastatus* (Kavalco *et al.*, 2013; Pazza *et al.*, 2018).

Em se tratando do complexo *A. bimaculatus* ao qual *A. altiparanae* pertence, *A. lacustris* e *A. assuncionensis* apresentam poucos sítios de As-51; *A. bimaculatus* não apresenta marcações e em *A. altiparanae* algumas poucas populações possuem sítios pequenos e escassos enquanto a maior parte das populações não possui nenhum (Kantek *et al.*, 2009; Vicari *et al.*, 2010; Kavalco *et al.*, 2011; Pazza *et al.*, 2018). O histórico de distribuição do As-51 discutido nos parágrafos anteriores, é consistente com a “hipótese da biblioteca” proposta por Fry e Salser (1977), em que uma mesma família de satélites pode seguir por caminhos evolutivos independentes nas diferentes espécies, isso inclui a abundância das sequências nos genomas e a distribuição dos sítios nos cromossomos (Utsunomia *et al.*, 2017), de modo que o As-51 teria então uma distribuição inicialmente restrita a poucos locais cromossômicos, sendo posteriormente disseminado pelo genoma (Pazza *et al.*, 2018).

Com relação ao nosso putativo AalSat24-51, a partir dos alinhamentos comparativos de sua sequência com a de outros satélites semelhantes (ApaSat01-51 e ApaSat15-51 de *A. paranae* (Duílio *et al.*, 2017) e o As-51 original de *A. scabripinnis* (Mestriner, 2000)), através do software MEGA X, observamos uma série de pares de bases compartilhados entre eles (ver figura 1): 22pb da sequência do nosso AalSat24-51 se alinham com a do ApaSat15-51 (semelhança de 43,13%). Além disso, o AalSat24-51 também possui 23pb semelhantes aos do As-51 original (45,09%) e 22pb semelhantes aos do ApaSat01-51 (43,13%).

O ApaSat15-51, possui 26pb de sua sequência alinhados à do As-51 original (semelhança de 50,98%) e 24pb alinhados à do ApaSat01-51 (47,05%). Por fim, o ApaSat01-51, que é o As-51 de *A. paranae* (Duílio *et al.*, 2017), tem seus últimos 32pb alinhados com os 32 primeiros do As-51 original (62,74% de semelhança), de modo que é possível observar uma provável transposição da porção final do As-51 original formando a parte inicial do ApaSat01-51, com duas mutações de diferença.

Dadas as altas taxas evolutivas dos DNAs satélites (Kuhn *et al.*, 2008), a semelhança do As-51 com elementos transponíveis (Mestriner *et al.*, 2000) e as similaridades dos alinhamentos apresentados no parágrafo anterior e em sua respectiva

figura 1, é possível supor que nosso putativo AalSat24-51, em conjunto com o ApaSat15-51 (Duílio *et al.*, 2017), sejam sequências novas derivadas do As-51 ou mesmo variantes desse satélite, nos quais alguns pares de bases foram rearranjados.

Como o TAREAN classifica o AalSat24-51 como tendo uma baixa Probabilidade do Satélite, ele provavelmente se encontra disperso pelo genoma, o que nos leva à fazer algumas proposições sobre a presença e ausência do As-51 nas diferentes populações de *A. altiparanae*: 1- Ou o As-51 está de fato ausente na maior parte das populações, ou não se encontra em tandem, estando tão disperso pelo genoma que o sinal não é forte o suficiente para ser detectado na FISH tradicional (como sugere a nossa baixa Probabilidade do Satélite para o seu semelhante, o putativo AalSat25-51). Uma situação parecida foi encontrada com o ApaSat20-18 e o ApaSat36-21, que não apresentaram sinal na FISH, mas tiveram sua presença genômica atestada em análises de bioinformática (Duílio, *et. al.*, 2017). 2- Outra possibilidade é que algumas das populações de *A. altiparanae* possuem sequências derivadas do As-51, como o possível caso do nosso AalSat24-51, nas quais as sondas desenhadas com base no As-51 original não mais hibridizam ou hibridizam parcialmente (o que poderia explicar as poucas populações que apresentam sítios pequenos e escassos na literatura (Kavalco *et al.*, 2011)). Algo similar já foi observado em *A. mexicanus*, que não apresentou marcações resolutas para o As-51, sugerindo um pequeno número de cópias ou uma homologia parcial entre alguns cromossomos e a sonda (Kavalco e Almeida-Toledo 2007).

Tabela 1: Resumo dos dados brutos oriundos do sequenciamento NGS de *A. altiparanae*.

Número de reads sequenciados:	33,665,935
Tamanho médio da sequência:	150 pb
Número total de bases sequenciadas:	5,050 Mb
Conteúdo G+C total:	38,32%

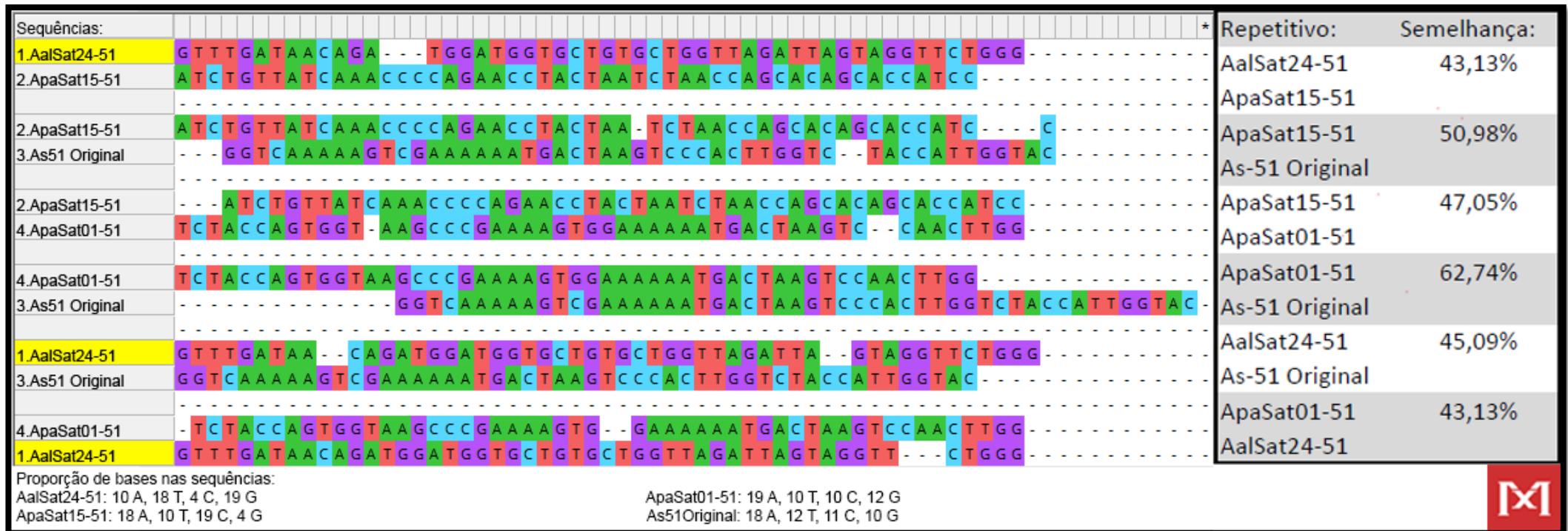
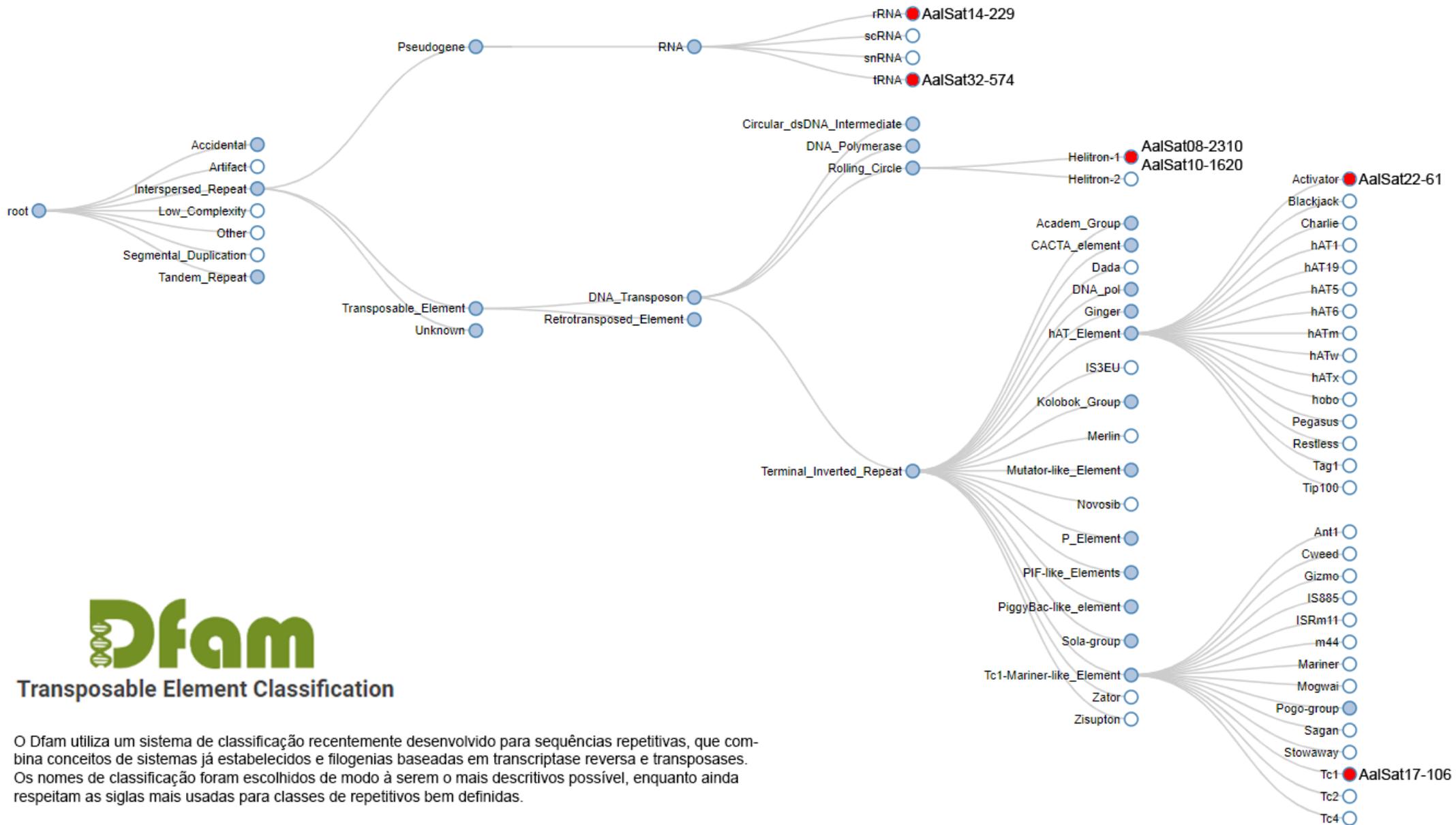


Figura 1: Alinhamentos da sequência do nosso satélite putativo AalSat24-51 (destacado em amarelo) com a dos outros satDNAs de 51pb de comprimento disponíveis para *Astyanax* na literatura. Os alinhamentos foram realizados individualmente par à par no software MEGA X, através do algoritmo Muscle nas configurações padrões. Em resumo no quadro à direita estão as semelhanças entre as sequências.

3.4.2. Classificação dos satélites putativos em relação ao banco de dados do Dfam:

Em se tratando do Dfam, o putativo AalSat31-574 se mostrou similar à pseudogenes repetitivos de tRNA-Thr-ACY e 4 dos outros possíveis satélites (AalSat08-2310, AalSat10-1620, AalSat22-61 e AalSat17-106) apresentaram semelhança não com DNAs satélites, mas sim com elementos transponíveis de diferentes subtipos (Figura 2). Elementos transponíveis são componentes importantes do genoma eucarioto e os peixes contêm mais famílias desses elementos em seu genoma do que outros vertebrados, como aves e mamíferos (Volf *et al.*, 2003; Fischer *et al.*, 2005; Zhou *et al.*, 2006). Essas sequências transponíveis desempenham um papel essencial como fonte de variação e evolução, seja induzindo mutações ou silenciando genes ao se deslocarem pelo DNA (Volf, 2005; Zhou *et al.*, 2006; Hickman *et al.*, 2010), Helitrons por exemplo, categoria na qual o AalSat08-2310 e o AalSat10-1620 foram classificados, podem atuar capturando e co-mobilizando genes e fragmentos de genes pelo genoma (Gupta *et al.*, 2005; Morgante *et al.*, 2005), isso também pode explicar o comprimento longo de suas sequências, que são umas das mais extensas dentre todos os repetitivos aqui encontrados.

Como os *A. altiparane* possuem um número diploide constante de $2n=50$ cromossomos e ainda assim apresentam variações em sua fórmula cariotípica e na simetria geral dos cariótipos (Kavalco *et al.*, 2011; Fernandes *et al.*, 2014), é possível que, por detrás desses pequenos rearranjos, sequências de elementos transponíveis estejam em atuação. Os repetitivos AalSat08-2310, AalSat10-1620, AalSat17-106 e AalSat22-61, foram considerados como diferentes tipos de elementos transponíveis. Tais elementos são diferenciados entre si em função de seus distintos mecanismos de transposição, mecanismos estes que podem ser diretamente ou indiretamente responsáveis por agrupar ou espalhar outras sequências pelo genoma sem que necessariamente ocorram rearranjos Robertsonianos.



Dfam Transposable Element Classification

O Dfam utiliza um sistema de classificação recentemente desenvolvido para seqüências repetitivas, que combina conceitos de sistemas já estabelecidos e filogenias baseadas em transcriptase reversa e transposases. Os nomes de classificação foram escolhidos de modo à serem o mais descritivos possível, enquanto ainda respeitam as siglas mais usadas para classes de repetitivos bem definidas.

Figura 2: Classificação das seqüências dos repetitivos AalSat14-229, AalSat32-574, AalSat08-2310, AalSat10-1620, AalSat22-61 e AalSat17-106 (destacados em vermelho) resultante da pesquisa realizada no banco de dados do Dfam.

DNA Satélite Putativo:	Confiança do Satélite:	Probabilidade do Satélite:	*Índice C:	*Índice P:	Tamanho consenso em pares de base:	Proporção no genoma (%):	Correspondência:	Banco de dados: Blast	Dfam
AalSat01-178	Alta	0,944	0,944	0,925	178	0,54	ApaSat10-179	75% 3e-50	.
AalSat02-91	Alta	0,767	0,979	0,886	91	0,35	.	.	.
AalSat03-151	Alta	0,944	0,992	0,933	151	0,32	.	.	.
AalSat04-26	Alta	0,989	0,988	0,951	26	0,016	.	.	.
AalSat05-35	Alta	0,78	0,975	0,881	35	0,016	.	.	.
AalSat06-43	Alta	0,761	1	0,9	43	0,11	.	.	.
AalSat07-487	Baixa	0,672	0,943	0,848	487	0,32	.	.	.
AalSat08-2310	Baixa	0,435	0,921	0,853	2310	0,32	Helitron-1N4_DR	.	3.4e-38
AalSat09-629	Baixa	0,579	0,934	0,866	629	0,3	.	.	.
AalSat10-1620	Baixa	0,0498	0,796	0,804	1620	0,27	Helitron-1N4_DR	.	6.7e-18
AalSat11-236	Baixa	0,0248	0,74	0,79	236	0,1	ApaSat04-233	86% 3e-51	.
AalSat12-3220	Baixa	0,0952	0,901	0,8	3220	0,099	.	.	.
AalSat13-189	Baixa	0,0626	0,859	0,821	189	0,094	ApaSat40-189	78% 4e-68	.
AalSat14-229	Baixa	0,232	0,913	0,897	229	0,055	rDNA 5S	98% 6e-78	1.1e-34
AalSat15-24	Baixa	0,028	0,752	0,758	24	0,052	.	.	.
AalSat16-185	Baixa	0,67	0,942	0,884	85	0,049	.	.	.
AalSat17-106	Baixa	0,0278	0,894	0,65	106	0,045	Tc1-4_DR	.	5.1e-06
AalSat18-40	Baixa	0,0388	0,97	0,728	40	0,039	Ame - mRNA variante X3	100% 7e-11	.
AalSat19-69	Baixa	0,0532	0,928	0,747	69	0,033	Ame - mRNA variante X1 e X2	95% 9e-20	.
AalSat20-22	Baixa	0,0303	0,774	0,624	22	0,033	ApaSat11-22	100% 0.027	.
AalSat21-42	Baixa	0,0292	0,916	0,667	42	0,031	.	.	.
AalSat22-61	Baixa	0,0291	0,838	0,626	61	0,03	hAT-N48_DR	.	3.4e-11
AalSat23-19	Baixa	0,0346	0,963	0,709	19	0,027	.	.	.
AalSat24-51	Baixa	0,0664	0,882	0,814	51	0,025	ApaSat15-51	75% 6e-9	.
AalSat25-112	Baixa	0,456	0,952	0,812	112	0,025	ApaSat33-112	83% 1e-36	.
AalSat26-58	Baixa	0,0204	0,721	0,584	58	0,024	.	.	.
AalSat27-76	Baixa	0,0187	0,735	0,526	76	0,024	.	.	.
AalSat28-85	Baixa	0,0181	0,769	0,529	85	0,016	.	.	.
AalSat29-24	Baixa	0,0563	0,833	0,8	24	0,014	.	.	.
AalSat30-142	Baixa	0,068	0,826	0,865	142	0,014	.	.	.
AalSat31-27	Baixa	0,0369	0,772	0,781	27	0,011	ApaSat25-27	85% 0.020	.
AalSat32-574	Baixa	0,612	1	0,889	574	0,01	tRNA-Thr-ACY	.	5.7e-25

Tabela 2: Satélites putativos encontrados em *A. altiparanae* através do RepeatExplorer e do TAREAN, seus valores de confiança e sua correspondência individual conforme os resultados obtidos no Blast e no Dfam. Os prefixos Apa e Ame remetem à *A. paranae* e *A. mexicanus*, respectivamente. No campo “Blast”, a porcentagem se refere ao *Query cover*, ao passo a função de “e” se refere ao *E-Value* em ambos os bancos de dados.

*Os índices C e P e sua importância estão explicados nos Materiais e Métodos e nos Resultados e Discussões.

3.5. Conclusões:

Através do sequenciamento de nova geração, da análise inteiramente *in silico* dos dados gerados e utilizando de uma abordagem relativamente recente via bioinformática com softwares robustos, porém, de fácil entendimento (RepeatExplorer e TAREAN), nós levantamos um total de 32 sequências significantes no satelitoma de *A. altiparanae*.

Minerando e comparando os dados genômicos do satelitoma de *A. altiparanae* com os de outras espécies do gênero, foi possível pontuar algumas hipóteses sobre a evolução cromossômica da espécie, mesmo sem a realização de hibridações fluorescentes *in situ*. As abordagens *in silico*, é claro, ainda não eliminam a necessidade da FISH para validação dos satélites putativos, mas fornecem não só um ponto de partida para investigações posteriores, como também sugere quais sequências são promissoras (como por exemplo os 17 satélites inéditos aqui apresentados), já que podem identificar até mesmo aquelas que não possuiriam sinal na hibridação.

Embora seja necessário validar os 32 repetitivos putativos através das técnicas de hibridação fluorescente *in situ*, ressaltamos que a organização genômica de alguns deles pode ser dispersa e não em tandem, o que dificulta sua identificação física nos cromossomos. De qualquer forma, todos os putativos levantados possuem potencial como marcadores citogenéticos, de modo que, a partir deste trabalho, apontamos o AalSat01-178, AalSat11-236 e o AalSat24-51 como mais promissores para *A. altiparanae*, em função de suas características aqui discutidas. Concluimos, portanto, que as novas abordagens citogenômicas são imprescindíveis para uma melhor compreensão sobre as sequências repetitivas e seu impacto na evolução do genoma e das espécies neotropicais, que como os *Astyanax*, são extremamente ricas e diversas.

3.6. Agradecimentos:

Os recursos computacionais para utilização da instância RepeatExplorer da plataforma Galaxy foram fornecidos pelo projeto ELIXIR-CZ (LM2015047), parte da infraestrutura internacional do ELIXIR.

3.7. Material Suplementar:

Figura 1: Workflow gerado na instância do RepeatExplorer dentro da plataforma Galaxy mostrando as etapas e ferramentas utilizadas para identificação *in silico* de DNAs satélite.

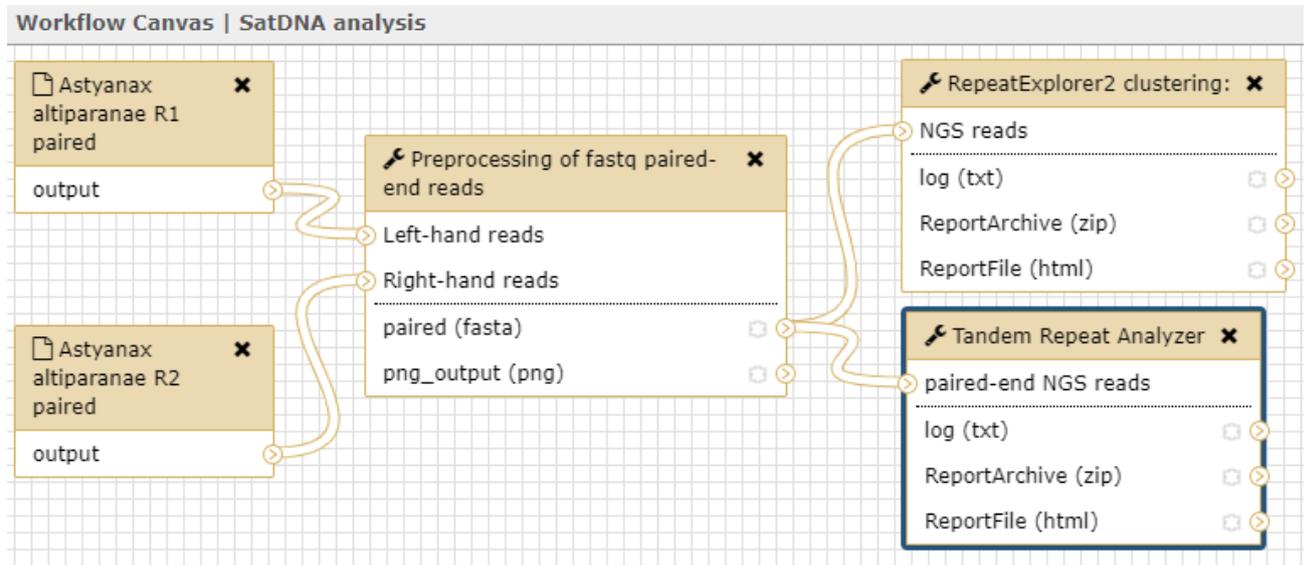


Tabela 1- Parâmetros do Preprocessing of fastq paired-end reads:

Tool Parameters

Input Parameter	Value	Note for rerun
Left-hand reads	1: Astyanax altiparanae R1 paired	
Right-hand reads	2: Astyanax altiparanae R2 paired	
Sequence sampling	false	
Quality cut-off	10	
percent above cutoff	95	
Trim sequences	false	
maximum Ns	0	
Do you want to use custom cutadapt options	false	
Use similarity search filtering	<i>not used (parameter was added after this job was run)</i>	
Rename sequences	False	

Tabela 2- Parâmetros do RepeatExplorer 1.0.0 e TAREAN:

Tool Parameters

Input Parameter	Value	Note for rerun
NGS reads	5: Interlaced paired reads from datasets 1 and 2	
paired-end reads	True	
Sample size	500000	
Select taxon and protein domain database version (REXdb)	Metazoa version 3.0	
Advanced options	false	
Select queue	basic_fast_queue	
Modify parameters (optional)	<code>-l select=1:ncpus=10:mem=32gb:scratch_local=50gb -l walltime=48:00:00 -q elixirre@pbs.elixir-czech.cz -v TAREAN_MAX_MEM=4000000,TAREAN_CPU=4</code>	

Referencial teórico das ferramentas de bioinformática:

Afgan E, Baker D, Van Den Beek M, Blankenberg D, Bouvier D, Čech M, ... & Grüning B: The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research*, 2016; 44-W1, W3-W10.

Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Čech M, ... & Guerler A: The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research*, 2018; 46-W1, W537-W544.

Andrews S: FastQC: a quality control tool for high throughput sequence data, 2010.

Bolger A M, Lohse M, Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data. In *Bioinformatics*, 2014; 30-15, pp. 2114-2120.

Novak P, Neumann P, Macas J: [Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data](#). *BMC Bioinformatics*, 2010; 11:378.

Novak P, Neumann P, Pech J, Steinhaisl J, Macas J: [RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads](#). *Bioinformatics* 2013; 29:792-793.

Novak P, Robledillo L A, Koblizkova A, Vrbova I, Neumann P, Macas J: [TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads](#). *Nucleic Acid Research*, 2017.

4. Considerações finais:

O gênero *Astyanax* como um todo, possui uma história evolutiva extremamente interessante e complexa, que é em partes refletida na diversidade cromossômica presente na maior parte de suas espécies. Algumas espécies, entretanto, mantêm arranjos cromossômicos relativamente conservados, com variações populacionais muito sutis, como no caso dos *A. altiparanae*, que demonstram que os métodos tradicionais da citogenética e da biologia molecular, apesar de detectarem essas variações, podem ser insuficientes para entendê-las.

Neste trabalho, testamos e utilizamos de uma série de métodos computacionais. Deste modo, exploramos via bioinformática os dados oriundos do genoma sequenciado de *Astyanax altiparanae*. Os métodos mais resolutivos, nos permitiram levantar uma série de dados e hipóteses que certamente contribuirão para a resolução do “quebra-cabeças” que é a filogenia dos *Astyanax*.

À começar pelo básico, que é a taxonomia do gênero, o genoma mitocondrial completo apresentado no capítulo 1 e as metodologias utilizadas para obtê-lo, viabilizam a confecção de diversos novos primers que podem ser úteis em estudos populacionais e para uma taxonomia mais integrativa, fornecendo também conhecimento que pode ajudar à identificar aquelas populações que, como unidades evolutivamente significativas, estejam seguindo caminhos evolutivos distintos.

A busca por entender como cada táxon se encaixa na história evolutiva dos *Astyanax*, se faz necessária porque seu processo compila uma gama enorme de conhecimentos sobre os mecanismos genéticos e cromossômicos por detrás de seus padrões evolutivos. Ao utilizar o mitogenoma completo de *A. altiparanae* para análises filogenéticas, obtivemos uma árvore filogenética cuja topologia corrobora a ideia de uma condição parafilética para o gênero, concordando também com filogenias anteriores baseadas na junção de dados moleculares e cromossômicos.

Neste cenário, a Bioinformática atual se insere não como substituta das técnicas laboratoriais clássicas, mas como um complemento imprescindível para elas. O avanço nas tecnologias de sequenciamento e no desenvolvimento de softwares acessíveis e robustos, está tornando a obtenção e a interpretação de dados genômicos muito mais intuitiva, ajudando à preencher algumas das limitações da citogenética e das técnicas de

FISH. No capítulo 2, por exemplo, descobrimos um total de 32 sequências repetitivas no genoma de *A. altiparanae* em estudos inteiramente *in silico*, ou seja, utilizando de ferramentas de bioinformática em conjunto com bancos de dados genéticos.

Por fim, ao comparar essas sequências com outras já disponíveis na literatura, pudemos apontar pelo menos três como marcadores citogenéticos promissores, constituindo um panorama sobre a composição do satelitoma de *A. altiparanae* e um ótimo ponto de partida para entender melhor os arranjos cromossômicos da espécie. Deste modo, a conjugação da bioinformática com as técnicas laboratoriais clássicas (citogenética e FISH) configuram abordagens citogenômicas modernas que, desde já, se mostram imprescindíveis para uma compreensão mais ampla sobre o funcionamento do genoma e dos mecanismos genético-cromossômicos por detrás da história evolutiva e da diversidade encontrada em *Astyanax*.

5. LISTA DE ABREVIÇÕES:

AalSat: DNA satélite de *Astyanax altiparanae*.

ApaSat: DNA satélite de *Astyanax paranae*.

Ala: Alanina.

Arg: Arginina.

ATPase: Adenosinatrifosfatase

Asn: Asparagina.

Asp: Ácido Aspártico.

Et al.: Abreviação de *Et alia*, do latim “e outros”.

EUA: Estados Unidos da América.

COI: Citocromo C oxidase subunidade I (I a III).

Cyt b: Citocromo B.

Cys: Cisteína.

DNA: Ácido desoxirribonucleico.

DNA C0t-1: DNA enriquecido com sequências altamente e moderadamente repetitivas.

DOP-PCR: reação em cadeia da polimerase iniciada com oligonucleotídeos degenerados.

FISH: Hibridação Fluorescente *in situ*.

Gln: Glutamina.

Glu: Ácido Glutâmico.

Gly: Glicina.

His: Histidina.

HTML: *HyperText Markup Language*, formato de dados.

Ile: Isoleucina.

LaGEEvo: Laboratório de Genética Ecológica e Evolutiva.

Leu: Leucina.

LTR: Repetições Terminais Longas das extremidades de retrotransposóns.

Lys: Lisina.

Met: Metionina.

mtDNA: DNA mitocondrial.

MG: Minas Gerais.

ND1 (1 a 6): NADH desidrogenase subunidade 1 (1 a 6).

NGS: *Next-Generation Sequencing* – Sequenciamento de nova geração.

NTS: Sequência espaçadora não transcrita.

Pb: Pares de bases.

PCGs: Genes codificadores de proteínas.

Phe: Fenilalanina.

Pro: Prolina.

rDNA: DNA ribossômico.

rRNA: RNA ribossômico.

RNA: Ácido ribonucleico.

satDNA: DNA satélite.

Ser: Serina.

SISGEN: Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado

SNPs: Polimorfismos de nucleotídeos únicos.

Thr: Treonina.

tRNA: RNA transportador.

Trp: Triptofano.

Tyr: Tirosina.

Val: Valina.

6. REFERÊNCIAS BIBLIOGRÁFICAS:

- Abel L D D S, Mantovani M, Moreira-Filho O: Chromosomal distribution of the As51 satellite DNA in two species complexes of the genus *Astyanax* (Pisces, Characidae). *Genetics and Molecular Biology*, 2006; 29.3, 448-452.
- Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, *et al.*: The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research*, 2018; 46.W1, W537-W544.
- Albert J S & Reis R E: *Historical Biogeography of Neotropical Freshwater Fishes*. Los Angeles: University of California Press, 2011; 408.
- Altschul S F, Gish W, Miller W, Myers E W, Lipman D J: Basic local alignment search tool. *J. Mol. Biol*, 1990; 215:403-410.
- Avise J C *et al.*: Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics, *Annu. Rev. Ecol. Syst.* 1987; 18, 489–522.
- Avise J C: Phylogeography: retrospect and prospect. *Journal of Biogeography*, 2009; 36.1, 3-15.
- Badyaev A V, Hill G E, Stoehr A M, Nolan PM, McGraw K J: The evolution of sexual size dimorphism in the house finch. II. Population divergence in relation to local selection. *Evolution*, 2000; 54.6, 2134-2144.
- Barreto C A V, Granja M M C, Vidigal P M P, Carmo A O, Dergam J A: Complete mitochondrial genome sequence of neotropical fish *Astyanax giton* Eigenmann 1908 (Ostariophysi; Characidae). *Mitochondrial DNA Part B*, 2017; 2.2, 839-840.
- Barbosa P, Leal E V, da Silva M, Almeida, M C, Moreira-Filho O, Artoni R F: Variability and evolutionary implications of repetitive DNA dynamics in genome of *Astyanax scabripinnis* (Teleostei, Characidae). *Comparative cytogenetics*, 2017; 11.1, 143.
- Betaco V A, Lucena C A: Two new species of *Astyanax* (Ostariophysi: Characiformes: Characidae) from eastern Brazil, with a synopsis of the *Astyanax scabripinnis* species complex. *Neotrop Ichthyol*, 2006; 4.1, 53-60.

Betancur-R R, Arcila D, Vari R P, Hughes L C, Oliveira C, Sabaj M H, Ortí G: Phylogenomic incongruence, hypothesis testing, and taxonomic sampling: The monophyly of characiform fishes. *Evolution*, 2019; 73.2, 329-345.

Boore J L: Animal mitochondrial genomes. *Nucleic acids research*, 1999; 27.8, 1767-1780.

Borowsky R: *Astyanax mexicanus*, the blind Mexican cave fish: a model for studies in development and morphology. *Cold Spring Harbor Protocols*, 2008; 11, pdb-emo107.

Bracewell R, Chatla K, Nalley M J, Bachtrog D: Dynamic turnover of centromeres drives karyotype evolution in *Drosophila*. *BioRxiv* [PrePrint]. 2019:733527. <https://doi.org/10.1101/733527> PMID: 31524597

Breda L, de Oliveira E F, Goulart E: Ecomorfologia de locomoção de peixes com enfoque para espécies neotropicais. *Acta Scientiarum. Biological Sciences*, 2008; 27.4:371-381.

Charlesworth B, Sniegowski P, Stephan W: The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, 1994; 371.6494, 215-220.

Curole J P & Kocher D T: Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends in Ecology & Evolution*, 1999; 14.10: 394-398.

Cohen E H, Bowman S C: Detection and location of three simple sequence DNAs in polytene chromosomes from *virilis* group species of *Drosophila*. *Chromosoma*, 1979; 73.3, 327-355.

de Freitas P D, Mendez F L, Chávez-Congrains K, Galetti P M, Coutinho L L, Pissinatti A, Bustamante C D: Next-Generation Sequencing of the Complete Mitochondrial Genome of the Endangered Species Black Lion Tamarin *Leontopithecus chrysopygus* (Primates) and Mitogenomic Phylogeny Focusing on the Callitrichidae Family. *G3: Genes, Genomes, Genetics*, 2018; 8.6, 1985-1991.

de Queiroz K: Species concepts and species delimitation. *Systematic biology*, 2007; 56.6, 879-886.

Domingues M S, Vicari M R, Abilhoa V, Wamser J P, Cestari M M, Bertollo L A C, de Almeida M C, Artoni F A: Cytogenetic and comparative morphology of two allopatric populations of *Astyanax altiparanae* Garutti & Britski, 2000 (Teleostei: Characidae) from upper rio Paraná basin. *Neotropical Ichthyology*, 2007; 5.1, 37-44.

Duílio M Z A S, Utsunomia R, Ruiz-Ruano F J, Daniel S N, Porto-Foresti F, Hashimoto D T, Oliveira C, Camacho J P M, Foresti F: High-throughput analysis unveils a highly shared satellite DNA library among three species of fish genus *Astyanax*. *Scientific reports*, 2017; 7.1, 12726.

Eigenmann C H: The american Characidae. 3. *Memoirs Museum of Comparative Zoology*, Harvard, vol. 43, no. 3, p.209-310 . Massachusetts, EUA, 1921.

Eigenmann C H: The american Characidae. 3. *Memoirs Museum of Comparative Zoology*, Harvard, vol. 43, no. 4, p. 311-428. Massachusetts, EUA, 1927.

Encyclopædia Britannica inc: *Astyanax*. Encyclopædia Britannica, 2018. Disponível em: <<https://www.britannica.com/topic/Astyanax-Greek-mythology>> Acesso em 18 de Novembro de 2019.

Eschmeyer W N & Fong J D: Catalog of Fishes, 2019. Disponível em: <<http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>> Acessado em 04/06/2019.

Fernandes C A, da Rocha R H, Bailly D, Guterres Z R, Alves D S, Martins-Santos, I C: Cytogenetic and Morphometric Analysis in the Species *Astyanax altiparanae* Garutti & Britski, 2000 (Teleostei, Characidae) from the Iguatemi River Basin, Brazil. *Folha biologica*, 2014; 62.3, 193-201.

Ferreira I A, Martins C: Physical chromosome mapping of repetitive DNA sequences in Nile tilapia *Oreochromis niloticus*: evidences for a differential distribution of repetitive elements in the sex chromosomes. *Micron* 39, 2008; 411–418. doi: 10.1016/j.micron.2007.02.010

Ferreira-Neto M, Artoni R F, Vicari M R, Moreira-Filho O, Camacho J P M, Bakkali M, Oliveira C, Foresti F: Three sympatric karyomorphs in the fish *Astyanax fasciatus* (Teleostei, Characidae) do not seem to hybridize in natural populations. *Comparative Cytogenetics* 2012; 6.1:29.

Ferretti A B S M, Ruiz-Ruano F J, Milani D, Loreto V, Martí D A, Ramos E, Martins C, Cabral-de-Mello D C: How dynamic could be the 45S rDNA cistron? An intriguing variability in a grasshopper species revealed by integration of chromosomal and genomic data. *Chromosoma*, 2019; 1-11.

Fischer C, Bouneau L, Coutanceau J, Weissenbach J, Ozouf-Costaz C, Volff J N: Diversity and clustered distribution of retrotransposable elements in the compact genome of the pufferfish *Tetraodon nigroviridis*. *Cytogenet Genome Res* 2005; 110:522–536.

Fowler H W: Os peixes de água doce do Brasil (1ª entrega). *Arquivo de Zoologia do Estado de São Paulo*, 1948; 6: 1–204.

Fry K & Salser W: Nucleotide sequences of HS- α satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell*, 1977; 12, 1069–1084.

Fu, Y T, Jin Y C, Liu G H: The Complete Mitochondrial Genome of the Caecal Fluke of Poultry, *Postharmostomum commutatum*, as the First Representative from the Superfamily Brachylaimoidea. *Frontiers in Genetics*, 2019; 10. doi:10.3389/fgene.2019.01037

Garavello J C & Sampaio F A A: Five new species of genus *Astyanax* Baird & Girard, 1854 from Rio Iguaçú, Paraná, Brazil (Ostariophysi, Characiformes, Characidae). *Brazilian Journal of Biology*, 2010; 70.3, 847-865.

Garutti V: Revisão taxonômica dos *Astyanax* (Pisces, Characidae), com mancha umeral ovalada e mancha no pedúnculo caudal, estendendo-se à extremidade dos raios caudais medianos, das bacias do Paraná, São Francisco e Amazônica (Tese de Livre-Docência não publicada), Universidade Estadual Paulista Júlio de Mesquita Filho, São José do Rio Preto, SP, 1995.

Garutti V, Britski H A: Descrição de uma nova espécie de *Astyanax* (Teleostei, Characidae) da bacia do alto Paraná e considerações sobre as demais espécies do gênero na bacia. *Comun. Mus. Cienc. Tecnol. PUCRS, Brazil. Serie Zool.* 2000; 3:65-88.

Gall J G, Cohen E H, Polan M L: Repetitive DNA sequences in *Drosophila*. *Chromosoma*, 1971; 33.3, 319-344.

GUERRA M: Introdução à Citogenética Geral. 1ª Edição, Editora Guanabara, 1988.

Goecks J, Nekrutenko A, Taylor J & Galaxy Team: Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology*, 2010; 11.8, R86.

Gupta S, Gallavotti A, Stryker G A, Schmidt R J, Lal S K: A novel class of Helitron-related transposable elements in maize contain portions of multiple pseudogenes. *Plant Mol Biol* 2005; 57:115–127.

Hashimoto D T, Ferguson-Smith M A, Rens W, Foresti F, Porto-Foresti F: Chromosome Mapping of H1 Histone and 5S rRNA Gene Clusters in Three Species of *Astyanax* (Teleostei, Characiformes). *Cytogenetic and Genome Research*, 2011; 134.1, 64–71. doi:10.1159/000323512

Hatanaka T & Galetti P M: Mapping of the 18S and 5S ribosomal RNA genes in the fish *Prochilodus argenteus* Agassiz, 1829 (Characiformes, Prochilodontidae). *Genetica*, 2004; 12.3, 239-244.

Helfman G S, Collette B B, Facey D E, Bowen B W: The diversity of fishes: Biology, Evolution and Ecology. John Wiley & Sons, Nova Jersey, EUA, 2009.

Heslop-Harrison J S: Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. *Plant Cell*. 2000; 12:617–35.

Hickman A B, Chandler M, Dyda F: Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Critical reviews in biochemistry and molecular biology*, 2010; 45.1, 50-69.

Homer, The Iliad with an English Translation by A.T. Murray, Ph.D. in two volumes. Cambridge, Massachusetts., Harvard University Press; London, William Heinemann, Ltd. 1924. Online version at the Perseus Digital Library access in Nov,18,2019 <http://data.perseus.org/citations/urn:cts:greekLit:tlg0012.tlg001.perseus-eng2:6.390-6.511>.

Hubley R, Finn R D, Clements J, Eddy S R, Jones T A, Bao W, Smit A F A, Wheeler T: The Dfam database of repetitive DNA families. *Nucleic acids research*, 2015; 44.D1, D81-D89.

Iwasaki W, Fukunaga T, Isagozawa R, Yamada K, Maeda Y, Satoh T O, Sado T, Mabuchi K, Takeshima H, Miya M, Nishida M: MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Molecular biology and evolution*, 2013; 30.11, 2531-2540.

Justi A J: Caracterização cariotípica de populações de *Astyanax fasciatus* (Cuvier, 1819) Pisces, Characidae, em três bacias hidrográficas (Unpublished Master's Thesis), Universidade Federal de São Carlos, São Carlos, SP, 1993; 83p.

Kantek D L Z, Vicari M R, Peres W A M, Cestari M M, Artoni R F, Bertollo L A C, Moreira-Filho O: Chromosomal location and distribution of As51 satellite DNA in five species of the genus *Astyanax* (Teleostei, Characidae, Incertae sedis). *Journal of Fish Biology*, 2009; 75.2, 408-421.

Kavalco K F & de Almeida-Toledo L F: Molecular cytogenetics of blind mexican tetra and comments on the karyotypic characteristics of genus *Astyanax* (Teleostei, Characidae). *Zebrafish*, 2007; 4.2, 103-111.

Kavalco K F, Pazza R, Brandão K O, Garcia C, Almeida-Toledo L F: Comparative cytogenetics and molecular phylogeography in the group *Astyanax altiparanae*–*Astyanax* aff. *bimaculatus* (Teleostei, Characidae). *Cytogenetic and Genome Research*, 2011; 134.2, 108-119.

Kavalco K F, Pazza R, de Oliveira Brandão K, de Almeida-Toledo L F. Biogeographic patterns in the chromosomal distribution of a satellite DNA in the banded tetra *Astyanax fasciatus* (Teleostei: Characiformes). *Organisms Diversity & Evolution*, 2013; 13.1, 67-76.

Kikuchi R, Yagi S, Kusuhara H, Imai S, Sugiyama Y, Kunio S: Genome-wide analysis of epigenetic signatures for kidney-specific transporters. *Kidney Int* 2010.

Koerber S: Freshwater fishes of Argentina: Etymologies of species names dedicated to persons. *Ichthyological Contributions of Peces Criollos*, 2011; 18, 1-18.

Kuhn G C S, Sene F M, Moreira-Filho O, Schwarzacher T, Heslop-Harrison J S: Sequence analysis, chromosomal distribution and long-range organization show that rapid turnover of new and old pBuM satellite DNA repeats leads to different patterns of variation in seven species of the *Drosophila buzzatii* cluster. *Chromosome Research*. 2008; 16.2:307–24. <https://doi.org/10.1007/s10577-007-1195-1>

Kuhn G C S, Küttler H, Moreira-Filho O, Heslop-Harrison J S: The 1.688 repetitive DNA of *Drosophila*: concerted evolution at different genomic scales and association with genes. *Mol. Biol. Evol.*, 2012; 29, 7–11.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K: MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, 2018; 35.6, 1547-1549.

Kursel L E & Malik H S: The cellular mechanisms and consequences of centromere drive. *Curr Opin Cell Biol*. 2018; 52:58–65. <https://doi.org/10.1016/j.ceb.2018.01.011> PMID: 2945425

Lima F C T, Malabarba L R, Buckup P A, Silva J F P, Vari R P, Benine R, Oyakawa O T, Pavanelli C S, Menezes N A, Lucena C A S, Malabarba M C S L, Lucena Z M S, Reis R E, Langeani F, Cassati L, Betarco V A, Moreira C, Lucinda P H F: *Genera incertae sedis* in Characidae. Check list of the freshwater fishes of South and Central America, 2003.

López-Flores I & Garrido-Ramos M A: The repetitive DNA content of eukaryotic genomes. In *Repetitive DNA* (Vol. 7, pp. 1-28). Karger Publishers, 2012.

Lower S S, McGurk M P, Clarck A G, Barbash D A: Satellite DNA evolution: old ideas, new approaches. *Current opinion in genetics & development*, 2018; 49, 70-78.

Lucena C A S, Menezes N A: Estudo filogenético da família Characidae com uma discussão dos grupos naturais propostos (Teleostei, Ostariophysi, Characiformes), 1993.

Mace G M: The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 2004; 359.1444, 711-719.

Malabarba L R & Weitzman S H: Descriptions of a new genus from southern Brazil, Uruguay and Argentina, with a discussion of a putative characid clade (Teleostei: Characiformes: Characidae). *Comunicacoes do Museu de Ciencias e Tecnologia da PUCRS*, 2003; 16.1, 67-151, 67, 151.

Martins C, Galetti P M: Chromosomal localization of 5S rDNA genes in *Leporinus fish* (Anostomidae, Characiformes). *Chromosome Research*, 1999; 7.5, 363-367.

Martins C & Galetti P M: Two 5S rDNA arrays in Neotropical fish species: is it a general rule for fishes? *Genetica*, 2001; 111.1-3, 439-446.

Mascagni F, Vangelist A, Giordani T, Cavallini A, Natali L: A computational comparative study of the repetitive DNA in the genus *Quercus* L. *Tree Genetics & Genomes*, 2020; 16.1,11.

Mehrotra S, Goyal V: Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. *Genomics, Proteomics e Bioinformatics*. 2014; doi:<https://doi.org/10.1016/j.gpb.2014.07.003>

Melo F A & Buckup P A: *Astyanax henseli*, a new name for *Tetragonopterus aeneus* Hensel, 1870 from southern Brazil (Teleostei: Characiformes). *Neotropical Ichthyology*, 2006; 4.1, 45-52.

Melters D P, Bradnam K R, Young H A, Telis N, May R M, Ruby J G, Sebra R, Peluso P, Eid J, Rank D, Garcia J F, DeRisi J L, Smith T, Tobias C, Ross-Ibarra J, Korf I, Chan S W L: Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome biology*, 2013; 14.1, R10.

Mestriner C A, Galetti Jr P M, Valentini S R, Ruiz I R G, Abel L D S, Moreira-Filho O, Camacho J P M: Structural and functional evidence that a B chromosome in the characid fish *Astyanax scabripinnis* is an isochromosome. *Heredity*, 2000; **85**, 1–9.

Mirande J M: Weighted parsimony phylogeny of the family Characidae (Teleostei: Characiformes). *Cladistics* 2009; 25.6: 574-613.

Mirande J M: Phylogeny of the family Characidae (Teleostei: Characiformes): from characters to taxonomy. *Neotropical Ichthyology* 2010; 8.3-385-568.

- Mirande J M: Morphology, molecules and the phylogeny of Characidae (Teleostei, Characiformes). *Cladistics*, 2018; 35.3, 282-300.
- Moreira-Filho O, Bertollo L A C: *Astyanax scabripinnis* (Pisces, Characidae): a species complex. *Genet Mol Bio* 1991; 14:331-357.
- Moscone E A, Samuel R, Schwarzacher T, Schweizer D & Pedrosa-Harand A: Complex rearrangements are involved in *Cephalanthera* (Orchidaceae) chromosome evolution. *Chromosome Research*, 2007; 15.7., 931-943.
- Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A: Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nat Genet* 2005; 37:997–1002.
- Nakatani M, Miya M, Mabuchi K, Saitoh K, Nishida M: Evolutionary history of Otophysi (Teleostei), a major clade of the modern freshwater fishes: Pangaeian origin and Mesozoic radiation. *BMC Evolutionary Biology* 11, 2011; Article number: 177.
- Nelson J S, Grande T C, Wilson M V H: *Fishes of the World*. John Wiley & Sons, Nova Jersey, EUA, 2016.
- Novák P, Neumann P, Macas J: [Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data](#). *BMC Bioinformatics*, 2010; 11:378.
- Novák P, Neumann P, Pech J, Steinhaisl J, Macas J: RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*, 2013; 29.6, 792-793.
- Novák P, *et al.*: [TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads](#). *Nucleic Acid Research*, 2017.
- Oliveira C, Avelino G S, Abe K T, Mariguela T C, Benine R C, Ortí G, Vari R P, Corrêa e Castro R M: Phylogenetic relationships within the speciose family Characidae (Teleostei: Ostariophysi: Characiformes) based on multilocus analysis and extensive in group sampling. *BMC Evolutionary Biology*, 2011; 11, 275–299.
- Orsi M L, Carvalho E D, Foresti F: Biologia populacional de *Astyanax altiparanae* Garutti & Brutti & Britski (Teleostei, Characidae) do médio Rio Paranapanema, Paraná, Brasil. *Revista Brasileira de Zoologia* 2004; 21.2-207-218.

Ornelas-Garcia C P, Dominguez-Dominguez O, Doadrio I: Evolutionary history of the fish genus *Astyanax* Baird & Girard (1854) (Actinopterygii, Characidae) in Mesoamerica reveals multiple morphological homoplasies. *BMC Evolutionary Biology*, 2008; 8, 340.

Padial J M, Miralles A, De la Riva I, Vences M: The integrative future of taxonomy. *Frontiers in zoology*, 2010; 7.1, 16.

Pamponét V C C, Souza M M, Silva G S, Micheli F, de Melo C A F, de Oliveira S G, Costa E A, Corrêa R X: Low coverage sequencing for repetitive DNA analysis in *Passiflora edulis* Sims: cytogenomic characterization of transposable elements and satellite DNA. *BMC genomics*, 2019; 20.1, 262.

Pavlek M, Gelfand Y, Plohl M, Meštrović N: Genome-wide analysis of tandem repeats in *Tribolium castaneum* genome reveals abundant and highly dynamic tandem repeat families with satellite DNA features in euchromatic chromosomal arms. *Dna Res.*, 2015; 22, 387–401.

Pansonato-Alves J C, Hilsdorf A W S, Utsunomia R, Silva D M Z A, Oliveira C, Foresti F: Chromosomal mapping of repetitive DNA and cytochrome C oxidase I sequence analysis reveal differentiation among sympatric samples of *Astyanax fasciatus* (Characiformes, Characidae). *Cytogenet Genome Res*, 2013; 141:133–142

Peleg S *et al.*: Altered histone acetylation is associated with age-dependent memory impairment in mice. *Science* 2010, 328:753-756.

Pazza R, Kavalco K F, Bertollo L A C: Chromosome polymorphism in *Astyanax fasciatus* (Teleostei, Characidae), 1. Karyotype analysis, Ag-NORs and mapping of the 18S and 5S ribosomal genes in sympatric karyotypes and their possible hybrid forms. *Cytogenetic and Genome Research*, 2006; 112.3-4:313-319.

Pazza R, Kavalco K F, Prioli S M A P, Prioli A J, Bertollo L A C: Chromosome polymorphism in *Astyanax fasciatus* (Teleostei, Characidae), Part 3: Analysis of the RAPD and ISSR molecular markers. *Biochemical Systematics and Ecology*, 2007; 35.12, 843-851.

Pazza R, Kavalco S A F, Penteadó P R, Kavalco K F, Almeida-Toledo L F: The species complex *Astyanax fasciatus* Cuvier (Teleostei Characiformes) – A multidisciplinary approach. *Journal of Fish Biology*, 2008b; 72: 2002–2010.

Pazza R, Argento R F, Mariano I H, Kavalco K F: Phenotypic plasticity and karyotype conservation in allopatric populations of *Astyanax* aff. *fasciatus* (Teleostei, Characiformes) at the Upper Paraná river basin. *Acta Zoologica*, 2016; 97.2, 241-245.

Pazza R, Jorge A D, Karine F K: Trends in Karyotype Evolution in *Astyanax* (Teleostei, Characiformes, Characidae): Insights From Molecular Data. *Frontiers in genetics*, 2018; 9: 131.

Peres W A M, Buckup P A, Kantek D L Z, Bertollo L A C, Moreira-Filho O: Chromosomal evidence of downstream dispersal of *Astyanax fasciatus* (Characiformes, Characidae) associated with river shed interconnection. *Genetica*, 2009; 137.3, 305.

Plohl M, Meštrović N, Mravinac B: Satellite DNA evolution. In *Repetitive DNA*, Karger Publishers, 2012; (Vol. 7, pp. 126-152).

Rebordinos L, Cross I, Merlo A: High evolutionary dynamism in 5S rDNA of fish: state of the art. *Cytogenetic and Genome Research*, 2013; 141.2-3, 103-113.

Richard G F, Kerrest A, Dujon B: Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.*, 2008; 72.4, 686-727.

Rodrigues P H D M, dos Santos R Z, Silva D M Z A, Goes C A G, Oliveira C, Foresti F, Porto-Foresti F, Utsunomia R: Chromosomal and Genomic Dynamics of Satellite DNAs in Characidae (Characiformes, Teleostei) Species. *Zebrafish*, 2019.

Rose H & March J: *Astyanax*. *Oxford Classical Dictionary*, 2015. <https://oxfordre.com/classics/view/10.1093/acrefore/9780199381135.001.0001/acrefore-9780199381135-e-889>. Acesso em 18 de Novembro de 2019.

Rosic S, Kohler F, Erhardt S: Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division (vol 207, pg 335, 2014). *J Cell Biol.* 2014; 207.5:673-.

Ruiz-Ruano F J, López-León M D, Cabrero J, Camacho J P M: High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Scientific reports*, 2016; 6, 28333.

Saito Y, Edpalina R R, Abe S: Isolation and characterization of salmonid telomeric and centromeric satellite DNA sequences. *Genetica*, 2007; 131: 157–166. doi: 10.1007/s10709-006-9126-1

Schmitter-Soto J J: A phylogeny of *Astyanax* (Characiformes: Characidae) in Central and North America. *Zootaxa*, 2016; 4109, 101–130.

Schwarzacher T: DNA, chromosomes, and in situ hybridization. *Genome*, 2003; 46.6, 953-962.

Shibatta O A, Artoni R F: Sobre a identidade das populações alopátricas de *Astyanax* (Characiformes, Characidae) das formações furna 1 e furna 2 do Parque Estadual de Vila Velha, Ponta Grossa, Paraná, Brasil. *Publication UEPG: Ciências Biológicas e da Saúde* 2005; 11.2.

Shibatta O A, Orsi M L, Bennemann S T, Silva-Souza A T: Diversidade e distribuição de peixes na bacia do rio Tibagi. Pp. 403-423. In: Medri, M. E., E. Bianchini, O. A. Shibatta & J. A. Pimenta (Eds.). *A bacia do rio Tibagi*. Londrina, 2002; 595p.

Silva D M D A, Utsunomia R, Ruiz-Ruano F J, Oliveira C, Foresti F: The complete mitochondrial genome sequence of *Astyanax paranae* (Teleostei: characiformes). *Mitochondrial DNA Part B*, 2016; 1.1, 586-587.

Silva B S, Heringer P, Dias G B, Svartman M, Kuhn G C: De novo identification of satellite DNAs in the sequenced genomes of *Drosophila virilis* and *D. americana* using the RepeatExplorer and TAREAN pipelines. *PloS one*, 2019; 14.12.

Smith D R: The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs?. *Briefings in functional genomics* 2015; 15.1: 47-54.

Struck T H, Feder J L, Bendiksbj M, Birkeland S, Cerca J, Gusarov V I, Kistenich S, Larsson K H, Liow L H, Nowak M D, Stedje B, Bachmann L, Dimitrov D: Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution*, 2018; 33.3, 153-163.

Tautz D: Notes on the definition and nomenclature of tandemly repetitive DNA sequences. In *DNA fingerprinting: State of the science* (pp. 21-28). Birkhäuser, Basel., 1993.

Telenius H K, Carter N P, Bebb C E, Nordenskjö M, Ponder B A J, Tunnacliffe A: Degenerate oligonucleotide primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics*, 1992; 13, 718–725.

- Treangen T J, Salzberg S L: Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* 2011, 13:36-46.
- Utsunomia R, Ruiz-Ruano F J, Silva D M Z A, Serrano E A, Rosa I F, Scudeler P E S, Hashimoto D T, Oliveira C, Camacho J P M, Foresti F: A Glimpse into the Satellite DNA Library in Characidae Fish (Teleostei, Characiformes). *Frontiers in Genetics*, 2017; 8.
- Utsunomia R, Silva D M Z A, Ruiz-Ruano F J, Goes C A G, Melo S, Ramos L P, Oliveira C, Porto-Foresti F, Foresti F, Hashimoto D T: Satellitome landscape analysis of *Megaleporinus macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. *Scientific reports*, 2019; 9.1, 5856.
- Vicari M R, Nogaroto V, Noletto R B, Cestari M M, Cioffi M B, Almeida M C, Moreira-Filho O, Bertollo L A C, Artoni R F: Satellite DNA and chromosomes in Neotropical fishes: methods, applications and perspectives. *Journal of Fish Biology*, 2010; 76.5, 1094-1116.
- Volff J N, Bouneau L, Ozouf-Costaz C, Fischer C: Diversity of retrotransposable elements in compact pufferfish genomes. *Trends Genet* 2003; 19:674-678.
- Volff J N: Genome evolution and biodiversity in teleost fish. *Heredity* 2005; 94:280-294.
- Zhou Q, Froschauer A, Schulthesis C, Schmidt C, Bienert G P, Wenning M, Dettai A, Volff J N: Helitron Transposons on the Sex Chromosomes of the Platyfish *Xiphophorus maculatus* and Their Evolution in Animal Genomes. *Zebrafish*, 2006; 3.1, 39-52.